Images, Numerical Analysis

of

Singularities and Shock Filters

Leonid Iakov Rudin

Computer Science Department
California Institute of Technology

5250:TR:87

Images, Numerical Analysis of Singularities and Shock Filters

Thesis by

Leonid Iakov Rudin

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

5250:TR:87

1987

Acknowledgments

I would like to thank my advisor James Kajiya for introducing me to various research areas ranging from Theory of Computation to Image processing and for his patience and support during my years at Caltech.

I am particularly indebted to Theo Pavlidis for memorable time spent at Bell Laboratories, Murray Hill and for initiating my search on the role of singularities in image processing. Illuminating discussions with him had a decisive influence on the contents of this thesis.

I would like to thank Tim Kay for his art of programming and help when I needed it the most.

I am grateful to members of my committee, Al Barr, Derek Fender, W.A.J. Luxemburg, Carver Mead, and Frederick Thompson, who taught me various courses and were instrumental in developing ideas for this dissertation.

I have special thanks to Jeffrey S. Schoenwald of Rockwell International for arranging my research at Science Center, where the idea of Shock Filters had crystalized.

I would like to thank Heinz-Otto Kreiss and Sukumar Chakravarthy for invaluable suggestions on the numerical methods for Shock Filters.

I would like to thank Zinovy Reichshtein for invaluable discussions on the distributional approach to the analysis of singularities.

I also thank my friends Chris Barrett and Young-Il Choo for many hours of insightful and stimulating discussions and years of comradeship. Many thanks to Devendra Kalra for helpful assisting in the jungle of the

Graphics Lab.

Finally, I would like to thank Nancy O'Connor who skillfully and patiently helped in the preparation of this manuscript.

Abstract

This work is concerned primarily with establishing a natural mathematical framework for the Numerical Analysis of Singularities, a term which we coined for this new evolving branch of *numerical* analysis.

The problem of analyzing singular behavior of nonsmooth functions is implicitly or explicitly ingrained in any successful attempt to extract information from images. The abundance of papers on the so called Edge Detection testifies to this statement.

We attempt to make a fresh start by reformulating this old problem in the rigorous context of the Theory of Generalized Functions of several variables with stress put on the computational aspects of essential singularities. We state and prove a variant of the Divergence Theorem for discontinuous functions which we call Fundamental Theorem of Edge Detection, for it is the backbone of the advocated here numerical analysis based on the estimates of contributions furnished by the essential singularities of functions.

We further extend this analysis to arbitrary order singularities by utilization of the Miranda's calculus of tangential derivatives. With this machinery we are able to explore computationally the internal geometry of singularities including singular, i.e., nonsmooth, singularity boundaries. This theory gives rise to singularity detection scheme called "rotating thin masks" which is applicable to arbitrary order n-dimensional essential singularities. In the particular implementation we combined first-order detector with derived here various curvature detectors. Preliminary experimental results are presented. We also derive a new class of nonlinear

singularity detection schemes based on tensor products of distributions.

Finally, a novel computational approach to the problem of image enhancement is presented. We call this construction the Shock Filters, since it is founded on the nonlinear PDE's whose solutions exhibit formation of discontinuous profiles, corresponding to shock waves in gas dynamics. An algorithm for experimental Shock Filter, based on the upwind finite difference scheme is presented and tested on the one and two dimensional data.

Contents

Dedicated to my parents

and Ohshima Sensei.

Images, Numerical Analysis of Singularities and Shock Filters

... for I had an eye
Which in my strongest workings, evermore
Was looking for the shades of difference
As they lie hid in all exterior forms,
Near or remote, minute or vast, an eye
Which from a stone, a tree, a wither'd leaf,
To the broad ocean and azure heavens,
Spangled with kindred multitudes of stars
Could find no surface where its power might sleep...[1]

[1] William Wordsworth, Prelude, Book III.

Chapter 1

Singularities of Images and Shock Filters

The visible contours of the objects are the projections of the bounding surface onto the retina of the eye. By examining the objects surrounding us, for instance, people's faces, we can study the singularities of visible contours. [1]

§1.1. *What do singularities of images tell the observer?*

1.1.1. *Is edge detection well-founded?.*

It is customary in the work related to edge detection to begin with examples where this procedure is applied within the framework of image analysis. And indeed applications are many!

Observe that rarely a paper on differential equations opens with an excuse of relevant utilization. Perhaps this last field is a few hundred years older and is accepted as a well-founded theory. While use should be the main motivating force, it cannot replace a firm theoretical foundation on which one can build. Yet, today a whole field of edge detection, and therefore related AI areas, are in need of mathematical foundation. This last statement might cause some legitimate protest. After all, literally hundreds of papers have been published in the last 30 years. With all due respect to the work done—and indeed a few important experimental results were accomplished, [Can.], [Hu.] [Na. Bi.]—try to find even the slightest

---

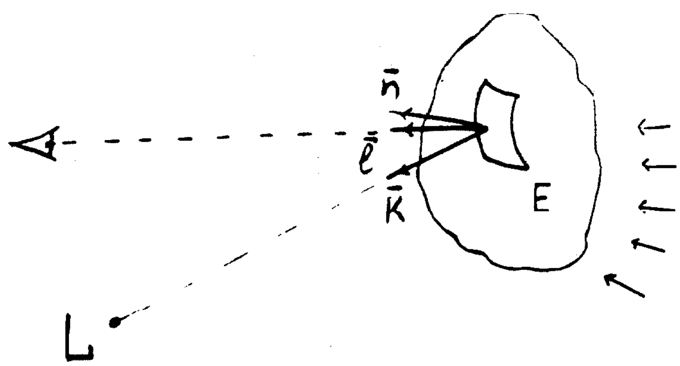
[1]Vladimir Arnold in Catastrophe Theory.

reference to this allegedly computing field in any modern text on numerical analysis. Even such a "cookbook" of computing as [P.F.T.V.] fails to do so. Or venture into the office of a numerical analyst, and read on his face this astonishment (and have second thoughts on the topic of one's thesis).

1.1.2. *Edge formation process.*

Thus I start this work with a brief look at the image formation process and why it is worthwhile at all to pay attention to singularities, which this process can produce. I shall also pay close attention to what kind of singularities arise in the real world pictures. Fortunately the topic of singularities of projection mappings has been a subject of several serious papers, such as [K.D.1.], whose 1976 paper took even geometers by surprise, and of earlier works on visual ecology [Gi.] and line drawings interpretation [KE.]. The last work provides an inventory of singularities evidently understood even by the cave man, as manifested by antipodal rock and cave artwork sources. The content of (1.1.3) is of an expository nature. My real intention is to draw readers' attention to possible deep connection with what I call a Numerical analysis of singularities. This will be the starting point of the investigation.

Let $F(x, y)$ be gray scale image, i.e., $F(x_0, y_0)$ is an absolute light intensity at the point (x_0, y_0) in the image plane, measured in the direction of point P on the observed surface as represented on Figure 1.1.



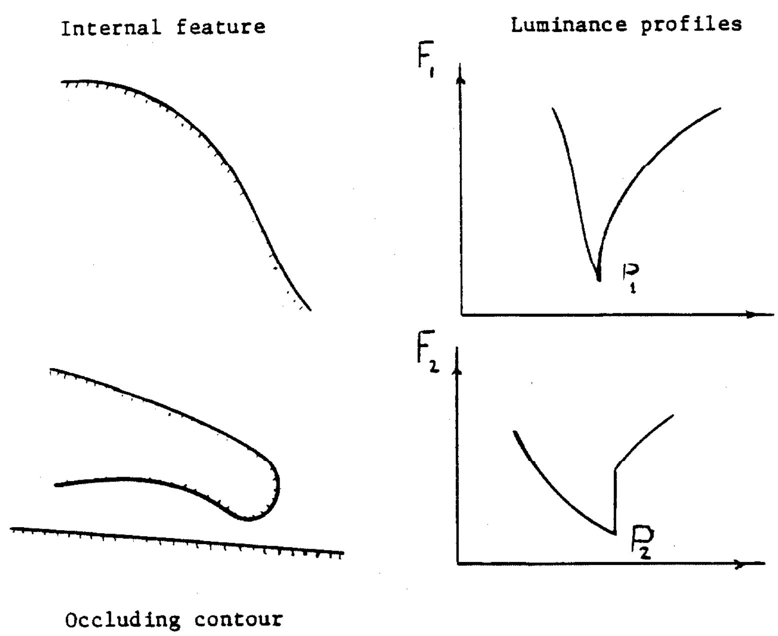
(Fig. 1.1) Diffuse illumination $\mathfrak{D}$.

Here E is a surface patch being observed, $\bar{n}$ is unit normal vector to E, $\bar{\ell}$- unit vector of the direction towards an observer, $\bar{k}$ is a unit vector in the direction of incident illumination, and L is a point light source whose direction is determined by the vector $\bar{k}$. The letters L and $\mathfrak{D}$ signify flux densities of the point and diffuse light sources respectively. Then the measured image intensity is given by

$$F(x, y) = \bar{n} \bullet \bar{k} \cdot L \cdot \phi_E + \mathfrak{D} \cdot \phi_E, \qquad (1.1)$$

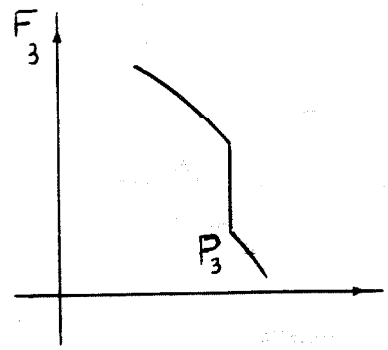
under the lambertian assumption, where ϕ_E is the reflectivity of the surface patch E. As normal vector $\bar{n}$ varies on the observed surface, its luminance will change, thus creating the shading effect. For detail on shape-from-shading see [Ho.], where continuous image intensities are exploited to determine three-dimensional surface shape. It has been postulated in [P.R.] that the perceptually important features occur at points where $\bar{n} \bullet \bar{k}$ is small, assuming that the camera and the light source are co-located. The reason for this is quite transparent, since $\bar{n} \bullet \bar{\ell} = 0$ means that adjacent points in the image plane may come from the points having nothing in common on the projecting surface. On the other hand, $\bar{n} \bullet \bar{k} = 0$ will define outlines of the shadow boundary on the surface. Of course, real pictures will have a combination of *extremal boundaries*, where the line of sight is *tangent* to the observed surface and boundaries caused by shadows.

While (from (1.1)) the outlines in the image will manifest themselves as discontinuities of the intensity function F(x, y), selfshadows will produce discontinuity in the first derivatives of F(x, y). Here by a selfshadow we mean a place on the surface with a smooth transition from full illumination to partial illumination by the diffuse light. It will be of critical importance for the treatment of singularities to determine the character of the luminance profiles across a feature. Simply speaking, I am very much interested in whether real images furnish the observer with the simple *jump-like* discontinuitities of the Heaviside function type, or a more complicated *singular* structure is present. Fortunately some experiments and calculations were performed in [Pe. Ro.] for several lighting models, including the single-source illumination, the mutual illumination and the diffuse illumination. Their findings can be summarized in Figure 1.2.



(Fig. 1.2)

And I would like to add to this an observation that the luminance profile at the boundary of a cast shadow in general appears as in Figure 1.3.



(Fig. 1.3)

Observe that $F_1(x, y)$ is a continuous function, but its *first derivative* is a discontinuous entity, while F_2 and F_3 are discontinuous functions.

Remark:

One cannot ignore that singularities can also be produced by markings on the surface, i.e., discontinuous reflectivity function ϕ in (1.1). We can safely assume that these will be of the simplest kind, i.e., step function. The art of camouflage is based on introducing this sort of markings for concealment and deception. Nature has used it to hide an animal in its environment [Lu.].

Assuming that P is an isolated singularity on the scan line, we may define one-sided limits of F(x) and its derivatives. Write them as $F_+^{(n)}(P)$ and $F_-^{(n)}(P)$ for $n = 0, 1, \ldots$.

We see that in general, the cast shadow will exhibit a "smooth" behavior across its jump discontinuity in the sense that

$$F^{(n)}_{3+}(P_3) = F^{(n)}_{3-}(P_3) \quad \text{for } n = 1, 2, \ldots .$$

However we have no such luck in the cases of internal features and occluding contours, i.e.,

$$F^{(n)}_{+}(P_2) \neq F^{(n)}_{-}(P_2), \; n = 0, 1, 2, \ldots$$

and

$$F^{(n)}_{+}(P_1) \neq F^{(n)}_{-}(P_1), \; n = 1, 2, \ldots .$$

Selfshadow behaves much like F_3 in terms of $F^{(n)}(P)$ for $n = 2, 3, \ldots$, i.e., higher derivatives.

Remark:

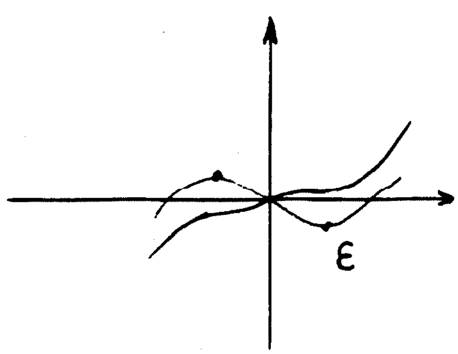
To summarize this observation we conclude that "interesting" features in the images not only may produce discontinuities, but also that singularities of a more complicated type than a simple jump can be formed.

1.1.3. *Invariant properties of the object as seen through the singularities of projection mapping.*

So far we have discussed essentially one-dimensional behavior of the image intensity function. We determined that points where this function is singular carry important information about the scene being observed. In particular, via geometry of the projection, orientation of the normal to surface at the occluding boundary can be found by computing the normal to detected edge which corresponds to this boundary.

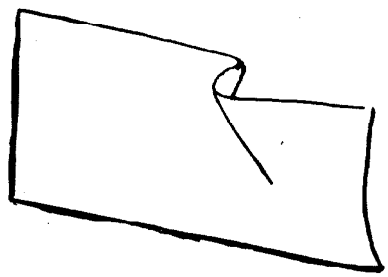
In their pioneering work [Ba. Te.] the authors call features of this kind *intrinsic*, since they are not a byproduct of the image formation process but, in fact, carry information of "what actually is there."

We know that some singularities in the *image function* correspond to the singularities of the *visual mapping*, which assigns to points of the surface a manifold of visual directions. The visual mapping is an orthogonal projection $p : R^3 \to R^2$, such that $p(x, y, z) = (x, y)$. If E is the observed surface, restrict p with $P_E : E \to R^2$. Singularities occur if the gradient of P_E vanishes, i.e., when P_E is *tangent* to E. For example, singularities of the projection of the sphere onto the plane are the points of the equator, which form a circle as an occluding boundary. H. Whitney in [Wh.] was able to classify completely all possible stable singularities of projections from two-dimensional manifolds (surfaces) to a two-dimensional manifold. Without going into the details of singularity theory, we say that a singularity is stable if it does not disappear after the projection or the surface are slightly perturbed, but the singularity just gets a little perturbed itself. For example, an occluding boundary does not go away if we shift around the object by a small amount. However, if a singularity's graph has a shape like in Figure 1.4, it may vanish under the small perturbution ϵ.



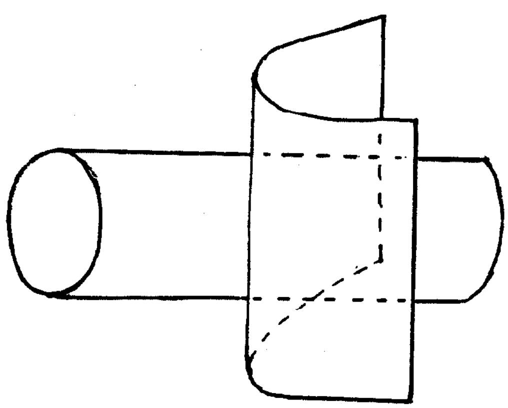
(Fig. 1.4)

Whitney proved that our projection function could have only two types of singularities, which he calls *folds* and *cusps*. The fold corresponds *exactly* to an occuding contour and the cusp happens to be a particular kind of internal feature, as on Figure 1.5.



(Fig. 1.5)

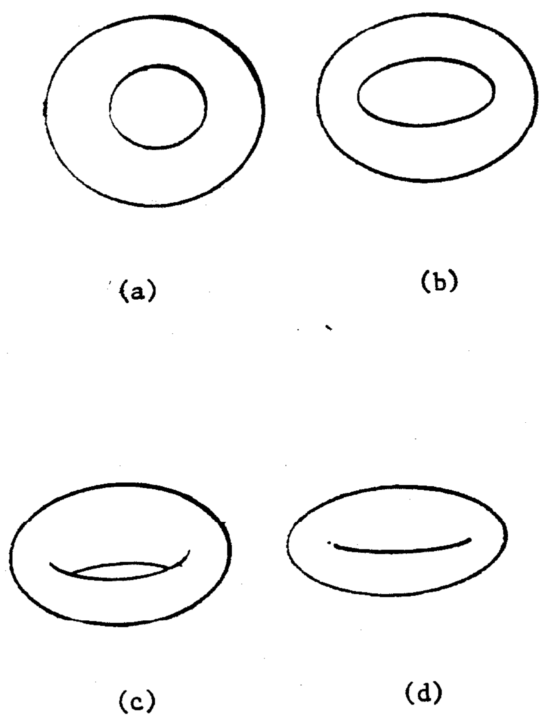
It is quite clear that features which vanish under small perturbations are of no interest at all in *seeing* them; that is, only *generic* surfaces shall be examined. If we allow a case of two surfaces being observed, then there is one more possibility, that of crossing (Fig. 1.6) or "junction."



(Fig. 1.6)

It turns out that if we change the projecting plane continuously, i.e., if an observer moves around, then stable singularities can undergo *metamorphoses* from one type to another, e.g., folds to cusps. Or they may disappear or be created. Concave edges may become convex. There is a profound relation between these events (called *catastrophes*) and the geometry of observed objects. In particular, *curvatures of the surface* relate to the *curvature of outlines*. In general, it can be shown [K. D.] that the *second fundamental form* of the surface will determine what kind of changes can occur in the observed singularities. This is a wonderful state of affairs. It says that by *observing just singularities of images, geometrical and topological (via junctions), properties of the object could be determined!* In such a way the shape of the object may be learned.

Figure 1.7 illustrates how rotation of a torus mutates its edges and simultaneously reveals its shape.



(Fig. 1.7)

(Pict. 1)

Note the *catastrophic* events on (c) and (d). While (a) and (b) could suggest that a cylinder with a hole is being observed, (c) and (d) destroy this hypothesis completely, suggesting at the same time that a doughnut is much more appropriate.

Picture sequence (Pict. 1) demonstrates rotation of the computer generated torus. Jump singularities are detected via application of the edge detector developed in the next chapter. The resulting *thresholded* edge map is superimposed with the orginal image. Notice that some of the prominent jump singularities are produced by the highlights.

In 1972 Waltz [Wa.] used an ad hoc classification of edges in an attempt to interpret line drawings. He produced a garden variety of labels which correspond to different types of edges and junctions and employed them in the Waltz filtering algorithm to converge to a single interpretation of the line drawings. Only one stationary line drawing was considered, thus no use of catastrophes. In fact, no attempt has been made as yet, to my best knowledge, to utilize singularity theory in the actual computation of shape from motion. All of the research seems to be concentrated on extracting information where it is *least likely to occur*, i.e., from smooth data points. Obviously this is not the way it is done in biological systems, because of the enormous complexity of the visual world around us.

Remark:

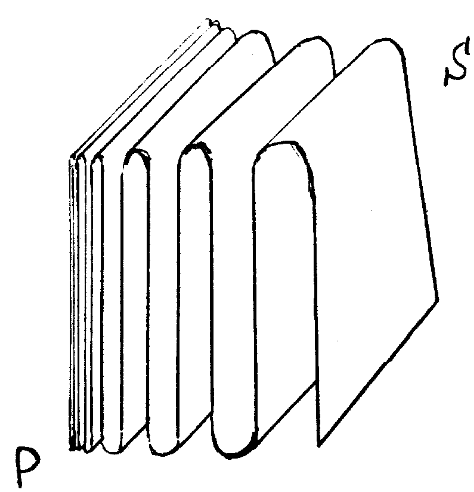
Observe that this use of singularities presupposes that their extraction from the image flow is a well defined and numerically sound process. For instance, to detect creation and annihilation of cusps and to infer changes of elliptic and hyperbolic intrusions, which happen suddenly, will demand powerful *curvature* detectors. These events are related to the observed surface being negatively or positively curved. Today, the arsenal of edge detection does not possess such a machinery.

1.1.4. *An Impossible Image.*

Finally I should ask a question: what kind of images are not possible? Paradoxically as it sounds, it will determine the mathematical model which we shall study here.

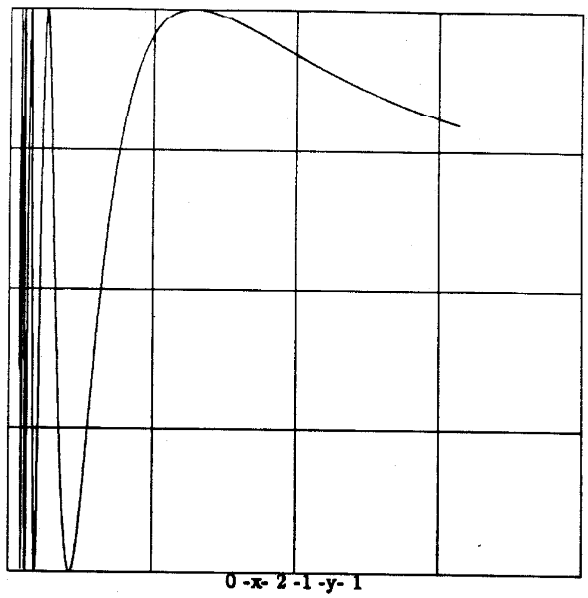
The answer will be similar to the response to how many angels can dance on the point of a needle.

An example of an impossible picture is presented in Figure 1.8.



(Fig. 1.8)

How does the image of the surface S appear to the observer near the point P? Since at P surface S has an infinite number of folds, i.e., an infinite number of occluding edges, nothing can actually be observed. Similarly one can produce a picture with an infinite density of cusps. But how about image intensity function F(x, y)? What kind of behavior would it exhibit per scan line? Figure 1.9 depicts this impossible picture.



(Fig. 1.9)

Does it remind us of anything from an introductory course of mathematical analysis? Of course, it is something like sin (1/x), a function whose *period* converges to <u>zero</u> at x = 0.

Look around you and see if there is anything like it in the real visual world. Put differently, could we possibly see it, even if it were there? Conclude the paragraph with the following postulate:

Postulate: Images belong to the space of functions of *bounded total variation.*

<u>**Moral:**</u> It is possible that a *myriad* of angels may dance on the point of

the needle, but who will *see* it?

§1.2. *Structure of singularities of functions of bounded total variation.*

The decision to narrow down to the space of functions of bounded total variation allows us to inquire as to the nature of the set of points where these functions are not smooth.

Generally speaking, images as we see them are the discrete measurements of the continuum world, which does not have a nice smooth behavior. Nonetheless, we are perfectly able to make a right judgment about what is there. In mathematical analysis, the functions which are equivalent *modulo Lebesgue measure zero* are not distinguished. If an image is a function F(x, y), its behavior on any set of 2-dimensional measure zero is an expendable entity. So why bother about edges? After all, intuitively they are just lines, whose 2-d measure (i.e., an area) is nil. To be able to reply, some measure theory should be introduced.

1.2.1. *Separation of regular and singular parts.*

Let μ be a Borel measure on R^n, then

Definition 1.1.

(i) $M = \{x \in R^n \mid \mu(x) \neq 0\}$ is called a set of pure points of measure μ

(ii) $\mu_{pp}(A) = \sum_{x \in A \cap M} \mu(\{x\}) = \mu(A \cap M)$ is called pure point measure of set A

(iii) $\mu_{cont} = \mu - \mu_{pp}$ is a continuous part of measure μ.

Then the following theorem is given by measure theory:

Theorem 1.1.

Any Borel measure is uniquely decomposed into the sum of pure point and continuous measures, i.e., $\mu = \mu_{pp} + \mu_{cont}$.

A few words of explanation. Usually measure takes a set and produces its "length, area," etc. If a set consists of just one point, its measure should be zero. Not so with a pure point measure; it assigns some "length" to some points which we call singular points.

Continuous part of the measure can be further decomposed in two parts:

Definition 1.2.

(i) μ is *absolutely continuous* with respect to Lebesgue measure λ if

$$d\mu = fd\lambda$$

for some $f \in L_{1_{loc}}(R^n)$.

(ii) μ is called *singular continuous* with respect to Lebesgue measure λ if there exists set A s.t. $\mu(A) = 0$ but $\lambda(R^n \setminus A) = 0$.

Absolutely continuous measures correspond therefore to the absolutely continuous functions, while singular continuous measure is much less obvious. The so called Devil's Staircase or Cantor's function is an example of a singular continuous measure. This is a curious case of the strictly monotonic continuous function whose derivative is equal to zero, a.e. This description of measure theory culminates with the Lebesgue decomposition theorem:

<u>**Theorem 1.2.**</u>

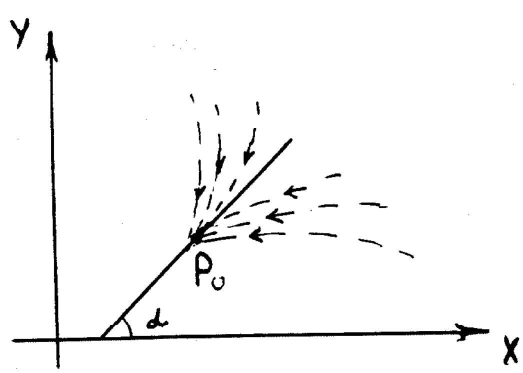
Any Borel measure μ can be uniquely decomposed into a sum

$$\mu = \mu_{pp} + \mu_{ac} + \mu_{sing.\ cont.} \tag{1.2}$$

If $F(x, y) \in L_{1_{loc}}$, i.e., $\iint_A |F(x, y)|\, dxdy$ exists for any compact set A, then F(x, y) is a Borel measure on R^2 and therefore is decomposable uniquely into the sum $F(x, y) = F_{pp}(x, y) + F_{ac}(x, y) + F_{sing.\ cont.}(x, y)$. For the time being we shall not worry about $F_{sing.\ cont.}(x, y)$. After all, its differential vanishes, and it is differentiation we are concerned with. In what follows, we will be mainly preoccupied with the structure of sets where singular parts of F are concentrated, i.e., analysis of $F_{pp}(x, y)$.

1.2.2. ***On how many singular points there are.***

Historically the first study of points of discontinuity of functions of several variables was undertaken by Kolmogorov and Verchenko in 1934 [Ve. Ko.]. Here is a brief summary of their results: Let $|F(P)| \leq M$ for all $P \in R^2$, i.e., it is uniformly bounded. Let $\bar{\phi}(P_0, \alpha)$ be a least upper bound of all $\overline{\lim}\,(F(P))$ when P approaches P_0 along any curve tangent to the direction α with axis x. See Figure 1.10.



(Fig. 1.10)

Similarly defined is the greatest lower bound $\underline{\phi}(P_0, \alpha)$. Then in order for $F(x, y)$ to be continuous in P_0, it is necessary and sufficient that

$$\overline{\phi}(P_0, \alpha) = \underline{\phi}(P_0, \alpha) = F(P_0).$$

Next, P_0 is a point of *discontinuity of a first kind* if $F(P)$ is not continuous at the point P_0 and for all angles α:

$$\overline{\phi}(P_0, \alpha) = \underline{\phi}(P_0, \alpha).$$

Theorem 1.3.

A set of points of discontinuities of the first kind is at most *countable*.

Further, P_0 is a *normal* point if for all α, both upper and lower limits are independent of the arcs from which they are approached.

$$\overline{\phi}(P_0, \alpha) = \overline{\phi}(P_0) \geq F(P_0) \geq \underline{\phi}(P_0) = \phi(P_0, \alpha), \tag{1.3}$$

Obviously every point of continuity is a normal point. Call normal points which are not points of continuity points of *normal jump*.

Theorem 1.4 (Kolmogorov).

A set of non-normal points can be put onto the countable number of rectifiable curves, and is of measure zero (here is the 2-dimensional Lebesgue measure).

Clearly, a set of points of *normal jump* is what we are after.

However Theorem 1.3 assures us that a set of points of the first kind is completely harmless, and that a set of non-normal points cannot be infinitely dense in the domain of F(x, y). These conclusions originate from the single assumption that the function is uniformly bounded. This condition, however, is not rich enough to make any judgments on the set of normal jumps, which corresponds exactly to the intuitive notion of an "edge." A. Volpert has continued this line of research in his 1967 paper [Vol.] on spaces of BV and quasilinear differential equations. He does not make any reference to [Va. Ko.] since his interest was solution of differential equations rather than real analysis of BV functions. His main concern was with assigning values to functions on the sets on which the *singular parts* of measures are concentrated. When nonlinear partial differential equations are written, it is common to have products and superpositions of discontinuous functions. Then one is interested in differentiation formulas for these combinations of functions, just like the conventional calculus rules. The following example demonstrates that usual calculus fails badly. Let h be a characteristic function of some set, then since $h^k = h$, we have

$$\begin{aligned} h' = (h^2)' = 2hh' = 4hh' - 2hh' = 4h^3h' - 2hh' \\ = (h^4 - h^2)' = (h - h)' = 0, \end{aligned} \tag{2.1}$$

which is obviously wrong, i.e., the rules we used do not apply to the discontinuous functions.

For the functions of one variable, total variation is defined as follows:

Definition 1.3.

Let f(x) be defined on [a, b]. It is said to have bounded variation if there exists such a constant K, that for any partition of [a, b] by points $x_0 = a < x_1 < x_2 \ldots < x_n = b$, the following inequality holds:

$$\text{Var } f(x) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq K. \tag{1.4}$$

The least of such K is called *total variation* of f(x) on [a, b], sup Var f(x) = TV(f(x)).

Several definitions could be given to extend this property to functions of several variables. They are associated with the names Vitali, Hardy, Frechet, Arzela, Pierpont and Tonelli. The most acceptable definition belongs to Tonelli:

Definition 1.4.

The function is assumed to be defined in the rectangle $a \leq x \leq b$, $c \leq y \leq d$. Denote $f_{x_0}(y)$ and $f_{y_0}(x)$ following functions of one variable:

$$f_{x_0}(y) = f(x_0, y),\ f_{y_0}(x) = f(x, y_0),$$

and let $TV_y(x_0)$, $TV_x(y_0)$ stand for total variations of $f_{x_0}(y)$ and $f_{y_0}(x)$ on the intervals [c, d] and [a, b] alone, as defined in Definition 1.3. Then put

$$\text{Var}_T = TV(f(x, y)) = \int_a^b TV_y(x_0)\, dx_0 + \int_c^d TV_x(y_0) dy_0. \tag{1.5}$$

If $\text{Var}_T < K$ for some constant K, f(x, y) is called to be of *bounded total* *variation in the sense of Tonelli.*

It is easy to see the meaning of Definition 1.4 for discrete grid functions. We simply compute the one-dimensional sum of jumps along each row and along each column and add them up to get TV(f(x, y)). For absolutely continuous functions of one variable, one sees that

$$\mathrm{TV}(f(x)) = \int_a^b |f'(x)|dx. \tag{1.6}$$

In the case of a continuously differential function of two variables, this formula could be extended to

$$\mathrm{TV}(f(x, y)) = \int_a^b\int_c^d |\mathbf{grad}\ f(x, y)|\ dxdy. \tag{1.7}$$

Properties of this variation were studied by Kronrod [Kr.] in great detail. Notice that while formula (1.5) essentially depends on the chosen coordinate system, formula (1.7) is invariant under coordinate transformations and thus could be extended to functions defined on arbitrary surfaces. Yet (1.5) has a nice geometrical interpretation as well, since it yields that surface $z = f(x, y)$ is of finite area if Var_T is bounded. It also has an additional merit of being a good candidate for a digital approximation of total variation for the discrete functions. In any case, either definition tries to capture the notion of the amount which the function f(x, y) *fluctuates* in the given rectangle. In such a way we may connect the integral of the modulus of gradient of an image to its informational content. We show in [Rud. 2] that this intuition could be made very precise by the so called ϵ-entropy concept.

Just as a function can be considered to be a measure, so are its derivatives.

Definition 1.5. [Fe.]

We say $f(x) \in BV(\Omega)$, where $x = (x_1, x_2, \ldots, x_n)$ if there exists a vector-valued set function μ, a Borel measure on Ω, such that

$$\int_{\Omega} \mathbf{grad}(\varphi) \cdot f dx = -\int_{\Omega} \varphi \, \mu(dx) \tag{1.8}$$

for any infinitely differentiable φ of compact support, and if it is locally integrable in $\Omega \subseteq R^n$.

This definition says that the space of functions of bounded total variation $BV(\Omega)$ consists of $L_{1_{loc}}$ elements, whose partial derivatives are measures. It says nothing about $f(x)$ being continuous; i.e., discontinuous functions are also members of $BV(\Omega)$. In what follows, μ is written as $\nabla f = \mathbf{grad}\ f$, because μ is a generalization of the notion of gradient. It may be shown that if $f \in BV(\Omega)$, then f is of bounded Tonelli variation. A typical example of a discontinuous function of bounded variation is

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} .$$

Then if

$$TV(\chi_A) = \int_{\Omega} |\nabla \chi_A| \, dxdy = P(A) < \infty, \tag{1.9}$$

P(A) is called a perimeter of set A and corresponds to the length of the boundary of the set A. Thus sets with very unusual boundaries can have a finite perimeter, which is measured by the total variation of its characteristic function. Similarly, for sets with finite perimeter the *inward* and *outward normals* could be defined, though in a much more involved way with the point set topology. The points where such a normal is defined constitute the so called *essential boundary* of the set A, called $\partial^* A$ [Fe.]. (Actually, the essential boundary has some other points, but their "length" is zero.)

In a way similar to Kolmogorov's (1.3) functions $\overline{\phi}$ and $\underline{\phi}$, Volpert [Vol.] introduces approximate limits $\ell_a(f(x_0))$ and $\ell_{-a}(f(x_0))$ when x_0 is approached along curves separated by the hyperplane $(\overline{x} - x_0, \overline{a}) > 0$, i.e., $\overline{a}$ being a normal direction of the hyperplane. Then x_0 is a *point of jump* if

$$l_a(f(x_0)) \neq -l_{-a}(f(x_0)).$$

If $\Gamma(f)$ is a set of points of jump for f and $x_0 \in \Gamma(f)$, then the *normal* to $\Gamma(f)$ at the point x_0 is *uniquely defined* for each point of jump x_0. The jump at the point x_0 is given by

$$|\Delta| f(x_0) = (l_a(f(x_0)) - l_{-a}(f(x_0)). \tag{1.9.1}$$

Further, it is shown that a set of non-normal points not only has "area" zero but also it has "length" zero, meaning that its contribution to the length of the sum of lengths of rectifiable curves on which it can be put (see Theorem 1.4) is zero. In precise terms, its (n — 1) dimensional

Hausdorff measure is zero. It also becomes possible to define the *mean* values for the BV functions, which exist not only up to n-dimensional measure zero but modulo (n − 1) dimensional measure zero—here is the essence of why boundaries which are sets of 2-dimensional measure zero cannot be assigned arbitrary values. The hidden reason for such exclusive behavior is that values on the boundary are actually determined by the conduct of the whole surrounding neighborhood of each singular point *on the boundary*. (This is not so with isolated singularities or regular values.) Boundaries then are locations of pure point measures, i.e., infinitely concentrated information.

Remark:

For a detailed account of how to assign values on the boundaries, consult any modern treatment of Traces of BV functions in the theory of minimal surfaces.

Mean values are produced by means of averaging kernels of the kind

$$g(x) = \begin{cases} \frac{2}{w_n}, & (x, a) > 0, \ |x| < 1 \\ 0, & \text{for all other } x \end{cases}, \tag{1.10}$$

i.e., an asymmetric spherical function where w_n is a volume of the n-dimension unit sphere.

Definition 1.6.

A set of class K is a Borel set, which is coverable by a finite or countable number of essential boundaries of sets with finite perimeter.

The following theorem provides us with an assurance that a set of jumps of $f \in BV$ may be safely confined to the number of rectifiable curves in its domain, therefore completing Theorem 4 of Kolmogorov.

<u>Theorem</u> <u>1.5</u> (Volpert).

Let $\Gamma(f)$ be the set of jumps of $f(x)$. Then there exists a set S of class K such that $\Gamma(f) \subseteq S \cup M$, where $H_{n-1}(M) = 0$.

For further applications in the edge detection we note that the mean value at x_0 is obtained by rotating the averaging kernel and integrating it with the function at x_0 until it reaches *maximum* for the unique vector $\bar{a}$. This vector $\bar{a}$ then defines direction of the *normal* to the set of jumps at x_0.

Finally, a generalization of Green's theorem

$$\int_{\Omega} f \cdot \operatorname{div} g \, dx = -\int_{\Omega} \langle g, \nabla f \rangle \, dx + \int_{\partial\Omega} f^{+} \cdot g \bullet \nu \cdot dH_{n-1}, \qquad (1.11)$$

where ν is a unit normal to $\partial\Omega$, $f \in BV(\Omega)$, and f^{+} is a trace on the boundary, $g_i \in C_{\infty}$ and $\operatorname{div} g = \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_i}$.

1.2.3. ***Synopsis.***

In the conclusion of this paragraph these most important points should be emphasized. They are the issues of vital significance when a problem of "computing the singularities" is at stake.

(i) Images, as we know them, can be modeled by the space of functions of locally bounded total variation $BV(\Omega)$.

(ii) The Lebesgue decomposition theorem yields that the singular parts of the functions should be uniquely separable from the smooth parts.

(iii) Performance of a BV function on its singular support (at the boundary) is *indispensable*, even though its n-dimensional measure is zero.

(iv) The notion of a *jump* is a well defined numerical quantity.

(v) A set of points of jump may be overlaid by a finite or countable number of rectifiable curves (genesis of edges).

(vi) A set of points which are neither jumps or points of continuity is very small (has *length* zero $H_{n-1}(S) = 0$).

(vii) Jump points are characterized by the *unique* vector, *normal* to the essential boundary which jump points comprise. At each point this direction is determinable by the limit of convolutions with the hemispherical kernel.

(viii) The generalized Green's theorem is valid.

Practical implications of this synopsis will manifest themselves in the next chapter.

§1.3. *Shock Filters*

In this section the notion of the Shock Filter, developed by the author in [Rud.1], is briefly introduced.

The aim of image enhancement is to preprocess an image into a form suitable for human or, what is a more important application, for machine analysis. This implies a redistribution of the original intensity flux into appropriate representation. As applies in particular to machine analysis, it is desirable to obtain a piecewise constant (staircase-like) representation for polyhedral edges or piecewise linear representation for surfaces with mutual illumination. In either case, discontinuity is present in the intensity function or in its derivative.

The computed representation shall contain all the esential information about the input image and should be free of spurious information.

Linear image filtering is the most common computational approach in achieving edge enhancement and detection. It can take various forms, such as spatial domain or spatial frequency domain; it can also have a probabilistic character. One way or another, all these forms can be reduced to the following equation:

$$G(x, y) = H(x, y) * F(x, y) \tag{1.12}$$

where G, H and F are the output image, space invariant filter, and original image, respectively.

The presence of noise should always be assumed, since even in apparently noiseless images noise is present in the form of quantization and numerical effects (e.g., rounding-off error).

For any filter H(x, y), the apriori positivity condition should hold for G(x, y) and F(x, y), since they are physical images. It is well known to every worker in the field, however, that this condition is hardly ever met. The class of filters that satisfies this apriori condition is that of smoothing filters. But this class is the least interesting for the purposes of *feature* extraction—e.g., edge extraction—because the smoothing linear filter by its very *linear* nature indiscriminately dissipates information from the image. Thus, not only is the noise removed, but also edges are blurred.

In general, the output of a non-smoothing linear filter is always plagued by the so-called *ringing* phenomenon, of which the *Gibbs oscillation* is a simple case. These spurious oscillations occur in proximity to edges and other singular features where all pictorial information is concentrated. Because it is impossible to distinguish these numerical artifacts from any actual information present in the image, further semantical processing is almost useless.

In this work, *ringing* has been rigorously defined in terms of certain functional behavior in order to capture all possible meanings of this elusive notion. The definition also implies that if filter is *ringing free* it always maintains positivity. Clearly it is natural to require following monotonicity property, which has been pioneered by A. Harten in his [Hart.] and has been recently the subject of intense studies in Numerical methods for scalar conservation laws known under the name TVD—*total variation* diminishing finite difference schemes.

<u>Definition 1.3.1</u>

Filter H(x, y) in (1.12) is said to be *monotonic* if no new local extrema in G(x, y) are created as compared to processed image F(x, y), and the local extremas present in F(x, y) are not accentuated.

This *information conservation* property is captured precisely by the relation

$$TV(G(x, y)) \leq TV(F(x, y)) \tag{1.13}$$

where G and F are given in (1.12) and TV is defined in (1.7). Following

[Hart.], we shall call these class filters—*total variation diminishing filters.*

It may be shown [Rud.2] that

<u>Lemma</u> <u>1.3.1</u> [Rud.2]

A linear filter is ringing free, i.e., TVD, if and only if it is a smoothing filter.

This lemma states that the linear systems theory is too weak to handle sophisticated information processing tasks, in particular, edge enhancement. Similar analysis of histogram modification and thresholding methods yields the same conclusion. Indeed, here the need appears for variable thresholding, indicating highly nonlinear algorithms.

Lemma 1.3.1 above is not very surprising either. The necessity for *smart* filters has been stressed by few authors [Pavl.], and some ad hoc nonlinear algorithms have been designed. Some of the most notable examples are homomorphic filters, directional filters, and what the author would like to classify as a class of hybrid filters and functional approximation filters.

Furthermore, Lemma 1.3.1 states more than just the need for nonlinearity. It pinpoints the exact mathematical nature of the limitation of linear, or equivalent Fourier transform, methods. Indeed, the mathematical notion of *ringing free* or total variation nonincreasing filters guides in the search for a well-founded nonlinear computational framework.

We proceed with a short description of the Shock Filter construction.

Intuitively, an edge is simply discontinous behavior of the analog image function and thus should be treated as such in any reasonable mathematical analysis of enhancement technique. Instead, all algorithms in this field are based on the rather vague notion of a digital edge (which is never rigorously defined) and thus hide the inherent weakness of this approach.

Hence our requirement is that discontinuity in the desired representation be analytical singularity rather than some digital version of it, even though eventually only a numerical approximation is computed. This is essential because only proper analytical consideration may yield proper numerical treatment, and not otherwise.

Until now no satisfactory discrete singularities theory has been developed in the context of multidimensional signal processing.

One can trace this apparent flaw of modern image processing in erroneous view that the only proper approximation space for digital signals is space B_ω of functions with Fourier spectrum concentrated on the interval $[-\omega, \omega]$. This implied approximation in L_2 norm. It is the opinion of the author that progress is to be achieved if distributional approach to singularities, developed in the next chapter, is applied in order to obtain Discrete Singularities Theory. In particular, space $BV(\Omega)$ of functions with bounded total variation should replace B_ω when digital approximations are considered.

Our search for appropriate ringing free enhancement filter should be limited only to operators which preserve non-negativity and locally conserve information. The intuitive sense of the locality property is to ensure that information is not globally spread over the image and that it does not *leak* across the *singularity boundaries* (see next chapter for the definition of this notion).

Begin with the study of quasi-linear partial differential equation of the form

$$a(t, x, u) \cdot u_t + b(t, x, u) \cdot u_x = c(t, x, u) \tag{1.14}$$

with two independent variables $t > 0$, $x \in R^1$, and coefficients a, b, c, being functions of t and x.

We consider the initial-value problem for (1.14), i.e., to find a solution of (1.14) for $t > 0$ if the given data is

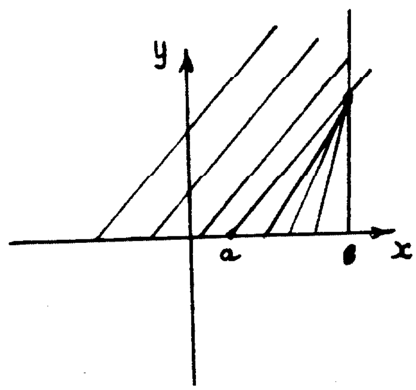
$$u(x, 0) = u_0(x), \quad x \in R_1 \tag{1.15}$$

The theory of equations of this type states that in general (1.14) cannot have a smooth solution defined for all $t > 0$, that is, the solution is not defined globally. This is true even if $u_0(x)$ is a smooth initial waveform.

The simplest example demonstrating this *singular behavior* is an equation of unidirectional nonlinear wave motion:

$$u_t + u \cdot u_x = 0 \tag{1.16}$$

The remarkable property of this equation is that it may lead to a discontinuous solution in finite time if the initial condition is not monotonically nondecreasing in space variable x, for each $t > 0$.



(Fig. 1.11)

Geometrically, this property is seen on Figure 1.11 which is characteristic diagram for the equation (1.16) when the initial condition is the right facing profile. Its characteristics represent trajectories in the phase space (x, t) along which information is carried from the initial curve $t = 0$.

In the case (1.16) the value of $u(x, t)$ is constant along each of the characteristic curves. If there are two points x_1 and x_2 such that $x_1 < x_2$ and $u(x_1, t) > u(x_2, t)$ the characteristic lines meet at some point *in* $t > 0$. Points on the initial wave with the larger u convects faster and overtake points on the wave convecting with the smaller u. At the point of a characteristic crossing the only reasonable physical counterpart to a multiple solution or *breaking* is to introduce a ***shock*** across which u changes discontinuously. In such a way, *analytical discontinuity* becomes a reality even for a smooth initial function.

There is much more to shock waves than is possible to discuss here.

But the notion of *entropy* is crucial for the intended application. It turns out that the solution to (1.16) is not unique precisely because of possible discontinuity. Then the only solutions not rejected are those that are *physically* meaningful.

A sufficient condition for this is the so called ***entropy criterion***:

> *The characteristics starting on either side of the discontinuity curve, when continued in the direction of increasing t, intersect the line of discontinuity.*

This shows that information is preserved locally, being compressed towards the shock point. Another consequence is that information does not cross the discontinuity point and therefore is not *leaked* across the boundaries.

Finally, the following condition is satisfied by the solutions of (1.16)

$$\mathrm{TV}(u(t_2, x)) \leq \mathrm{TV}(u(t_1, x)) \quad \text{if } t_2 \geq t_1, \tag{1.17}$$

and positivity is preserved.

Consider linear diffusion equation

$$u_t = \epsilon \cdot u_{xx} \tag{1.18}$$

with the initial condition $u(x, 0) = u_0(x)$. This physical process *dissipates information* across singularity boundaries in the smoothest possible fashion, i.e., it convolves the initial waveform with Green's function, which is a Gaussian. Thus it is a low pass filter (using systems theory jargon).

The simplest interaction of the nonlinear convection and dissipation is Burger's equation:

$$u_t + u \cdot u_x = \epsilon \cdot u_{xx} \tag{1.19}$$

It represents a balance between two physical processes: convection and diffusion. Here convection tends to steepen the gradients of the initial profile and diffusion does just the opposite. The diffusion is constant ϵ controls the magnitude of the spread. Controlling this parameter an interesting balance could be achieved: the magnitude of u is somewhat decaying, but right facing profiles get steeper and steeper. Thus any high frequency signal with small amplitude relative to $u_o(x)$ (e.g., *noise*) and riding on top of it will be eventually dissipated and concurrently *some* of $u(x)$ profiles get arbitrary sharp, while others are smoothed.

It is worthwhile to note that all the desired properties of Lemma 1.3.1 and (1.16) are still valid for (1.19).

However, in using (1.19) as a prototype filter other difficulties need to be overcome:

(i) First (1.19) is not symmetric in x-space. Indeed, it will *spread* all the left facing profiles.

(ii) Second, the solution is not stationary but propagates in space with different parts of the initial waveform traveling with different velocities.

(iii) Third, (1.19) is a one-dimensional equation, while a *two-dimensional filter* is required for image enhancement.

To resolve the first obstacle we modify the equation (1.19) to the space symmetric form

$$u_t - |u_x| \cdot u = \epsilon u_{xx} \tag{1.20}$$

Solutions of this equation have the property of forming *shocks* from both left and right facing profiles, as we will demonstrate numerically.

The unwanted *propagation* problem is alleviated by further transforming the equation into the form

$$u_t - f(u_x) \cdot |u_x| \cdot u = \epsilon u_{xx} \tag{1.21}$$

where coefficient $f(u_x)$ is set effectively to control *local* speed of the wave propagation in such a way that it monotonically goes to zero once some derivative bound is exceeded. In the digital implementation below its interpretation is the maximal size of the jump which is allowed to propagate after it has been formed. Thus it corresponds to the lower bound of the desired *image scale*. Every detail, whose size is smaller then this bound will eventually disappear from the image. Since speed of wave propagation in the finite difference schemes depends on the ratio number $b \cdot \frac{\Delta t}{\Delta x} = CFL$ where b is velocity coefficient in (1.14), and CFL stands for Courant, Friedricks and Lewy, who first studied the effect of this ratio on the stability of the finite difference schemes. Clearly if we make CFL to depend locally on $u'(x)$, we may then turn the wave propagation on and off in the digital implementation. Thus $f(u_x)$ is *local CFL number*.

The third problem, of extending (1.21) to the multidimensional case is the most difficult from the theoretical and numerical viewpoint. For example, a straightforward extension to the *two dimensional* Burger's equation

$$u_t + uu_x + uu_y = 0 \tag{1.22}$$

is unacceptable since application of the *rotation* transformation

$$
\begin{aligned}
x' &= (x + y)/\sqrt{2} \\
y' &= (y - x)/\sqrt{2}
\end{aligned}
\tag{1.23}
$$

the equation (1.22) is transformed into the *unidimensional* Burger's equation:

$$u_t + \sqrt{2} \cdot uu_{x'} = 0 \tag{1.24}$$

which will *blur* any singularity which is not normal to the axis x′. To put it differently, we need rotation invariant Shock Filter, as we shall call this class of function transformations.

This is done by considering differential equation in the form

$$u_t - F(|\nabla u|) \cdot u = \epsilon \cdot \Delta u \tag{1.25}$$

since the magnitude of the gradient of u is rotation invariant quantity.

Finally we solve (1.25) by the method of fractional steps, when half of the time the wave is propagating along the x-axis and second half along the y-axis. This scheme is typically applied to multidimensional finite difference equations. To solve (1.25) we have chosen one of the simple stable numerical methods available in the rich arsenal of today's finite difference methods for nonlinear conservation laws. We have used here a modified first order upwind Flux Correction Method of [Bor. Book].

This first order upwind scheme as modified by the author intricate switching mechanism which assures that information in the waveform is not propagated across image features, i.e., only *in the direction of the wind*. The algorithm is programmed in *divergence form*, i.e.

$$u_i^{n+1} = u_i^n - \mathrm{sign}(u_{i-1}^n - u_{i+1}^n) \cdot CFL(u^n, i) \cdot (Flux(u^n, i+1) - Flux(u^n, i)) + \\ + K \cdot (u_{i+1}^n - 2 \cdot u_i^n + u_{i-1}^n) \tag{1.26}$$

Where CFL is local Courant number and Flux is a numerical propagation image flux computed according to the direction of local characteristic speed w. The numerical flux is computed as in

$$Flux(u, i) = \begin{cases} F_{i-1}, & w_{i-1/2} \geq 0 \\ F_i, & \text{if } w_{i-1/2} < 0 \end{cases} \tag{1.27}$$

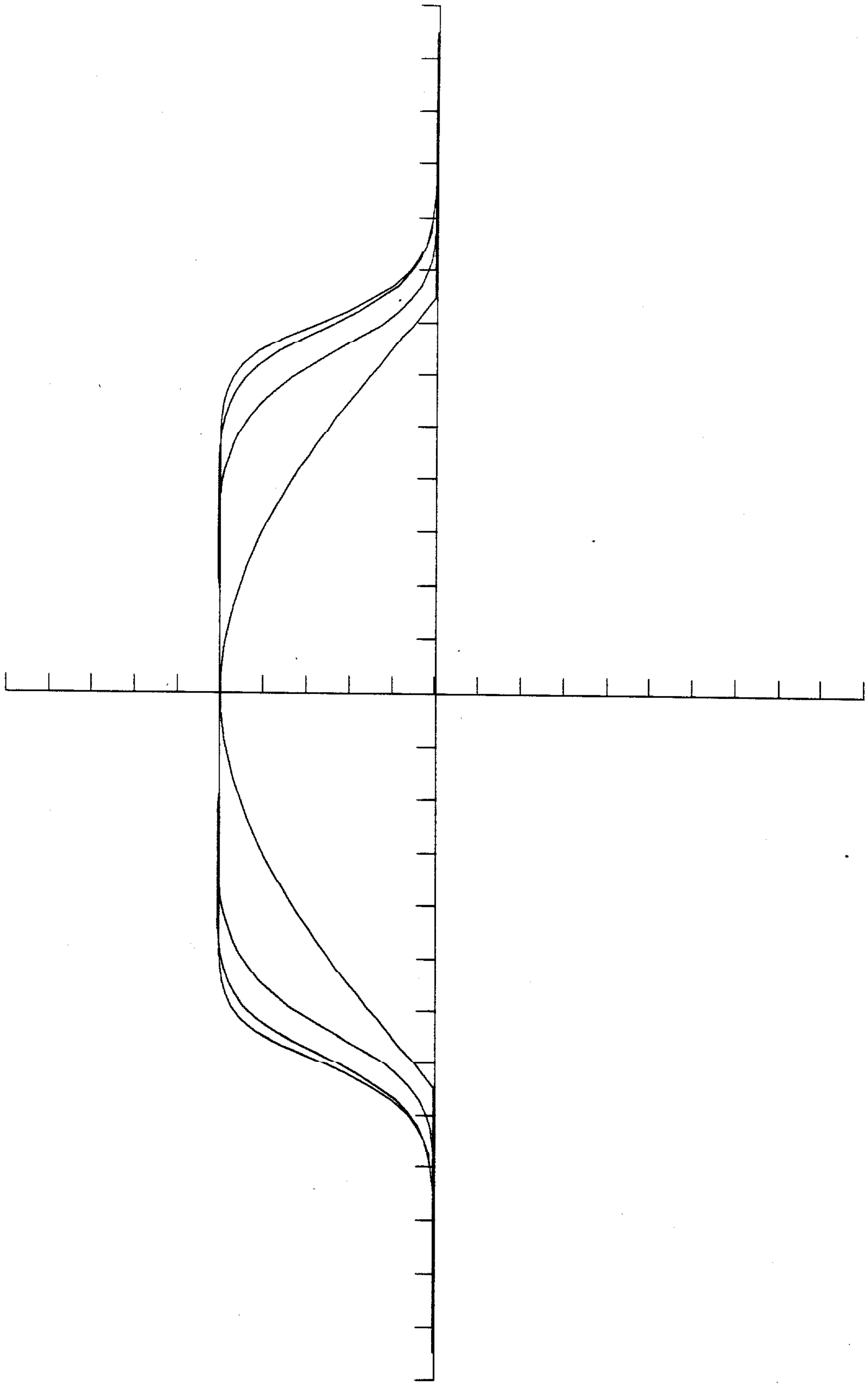
where $w_{i-1/2} = -(F_i - F_{i-1})$ and $F_i = \frac{1}{2} \cdot u_i^2$.

The C code of this finite difference Shock Filter is included. When this scheme was applied to various initial wave forms, interesting enhancement effects were obtained as demonstrated by Plot 1–Plot 7. In particular, an additive Gaussian white noise on Plot 3 has been successfully eliminated as well as uniform noise on Plots 4, 5, while the edge was preserved.

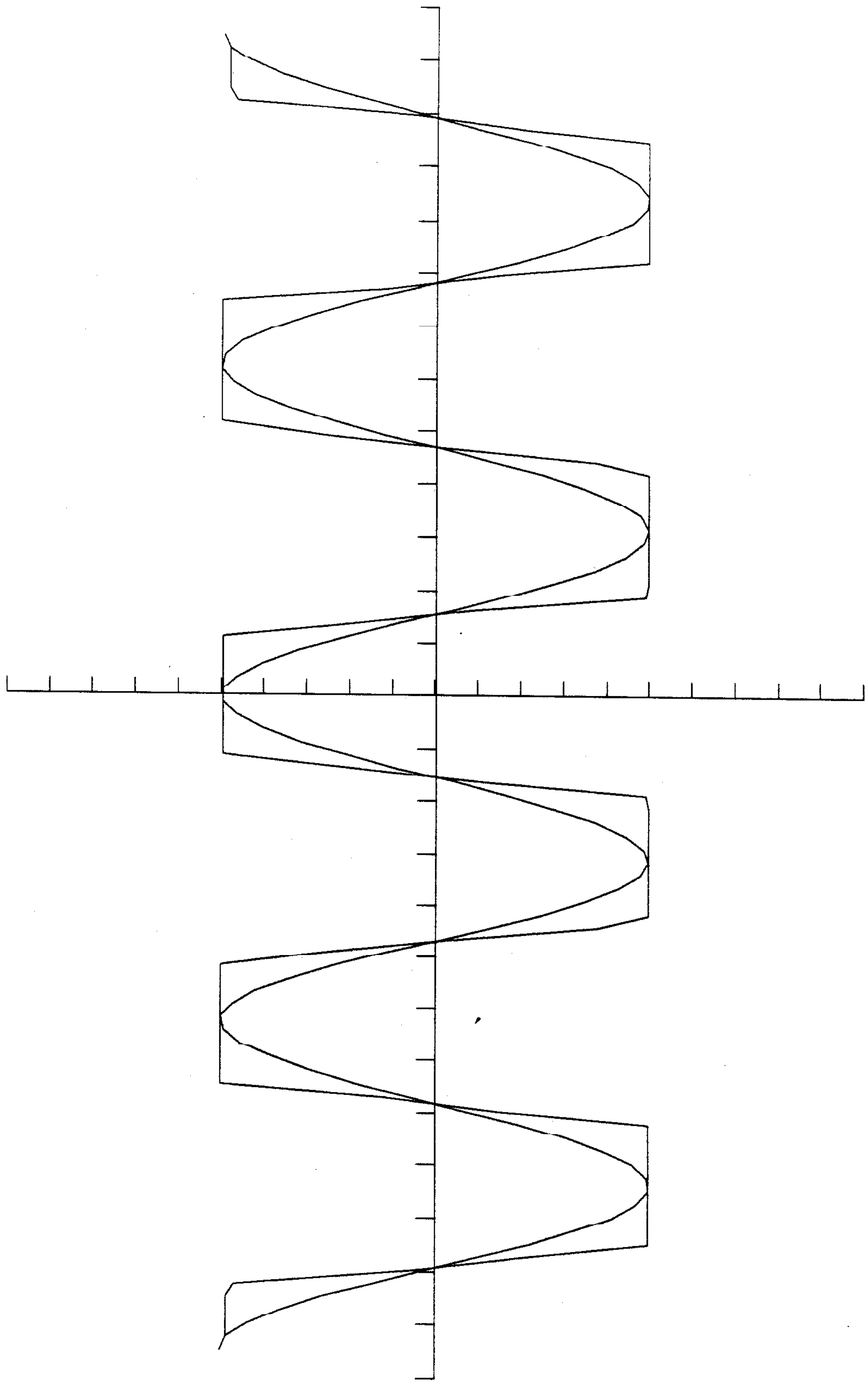
Plots 5 and 6 demonstrate how the Shock Filter continuously eliminates hierarchy of scales in succession, so that at the end only the highest scale–the "bump" is preserved.

We also have preliminary results with two-dimensional fractional step implementation of the above scheme, which is shown as the "Clock 1" sequence. Clearly these experiments are to continue with much sharper TVD methods and truly multidimensional difference schemes.

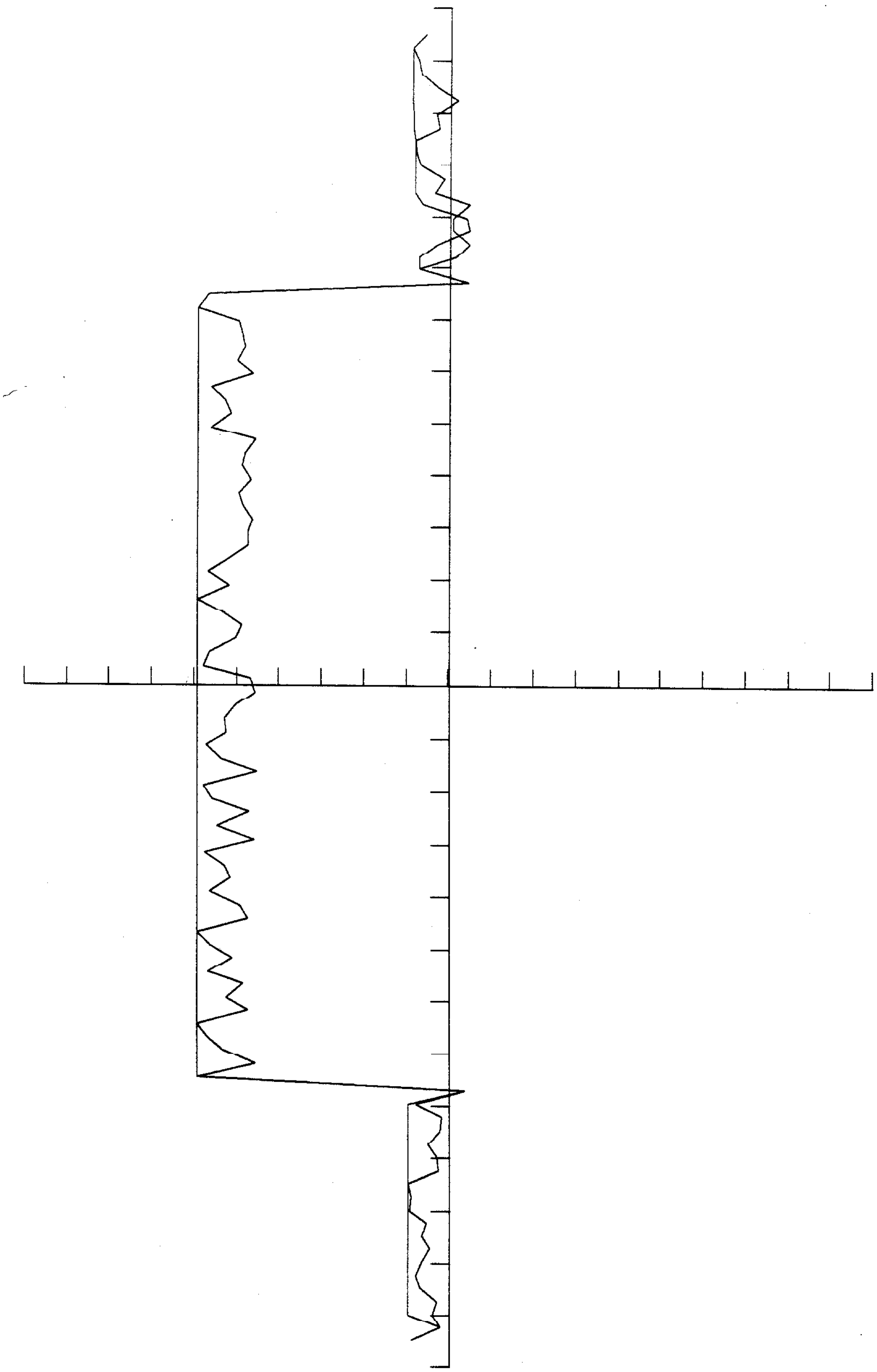
However, our computation has proved that it is possible to achieve *edge enhancement and simultaneous smoothing of high frequency noise.*



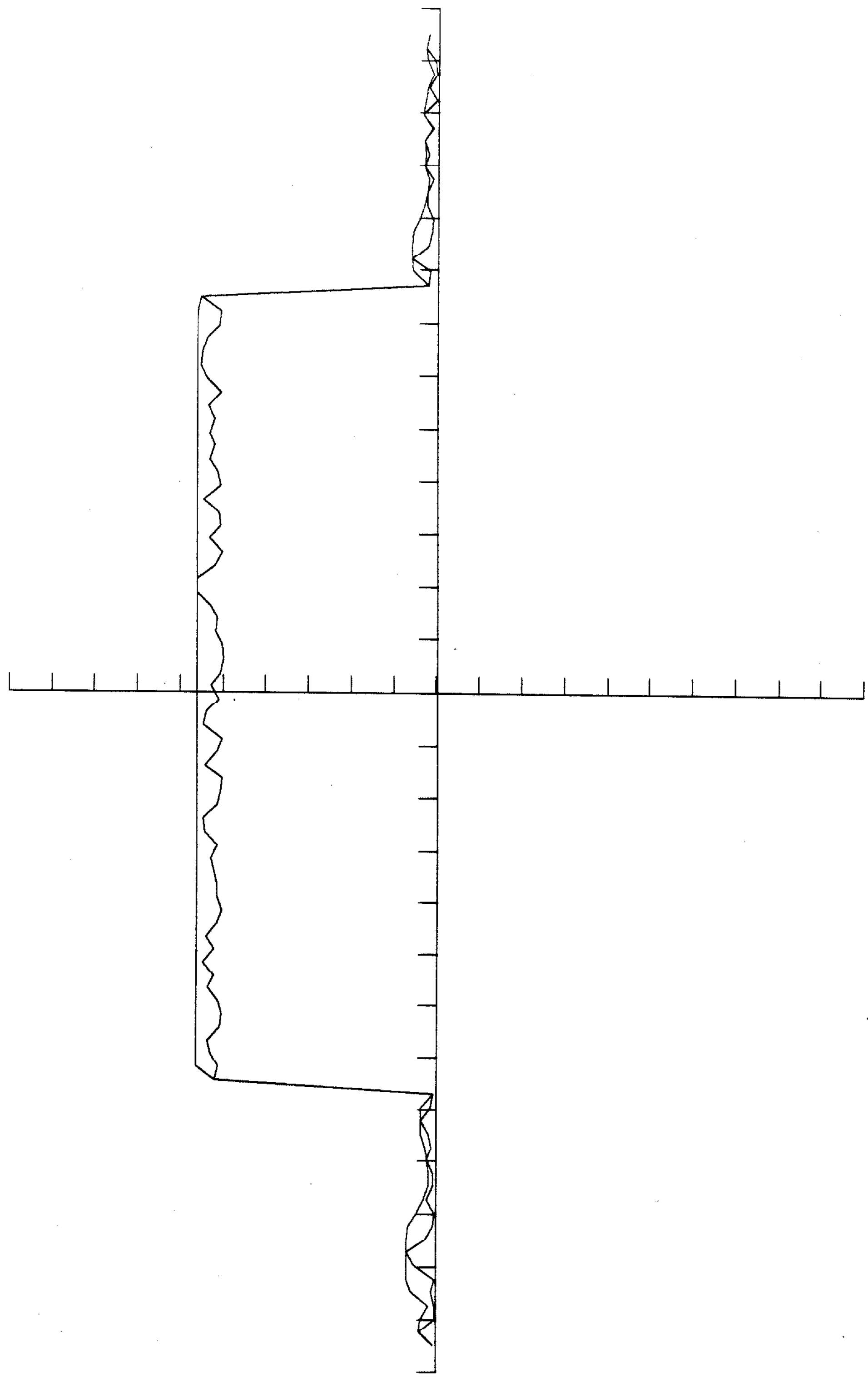
Plot 1



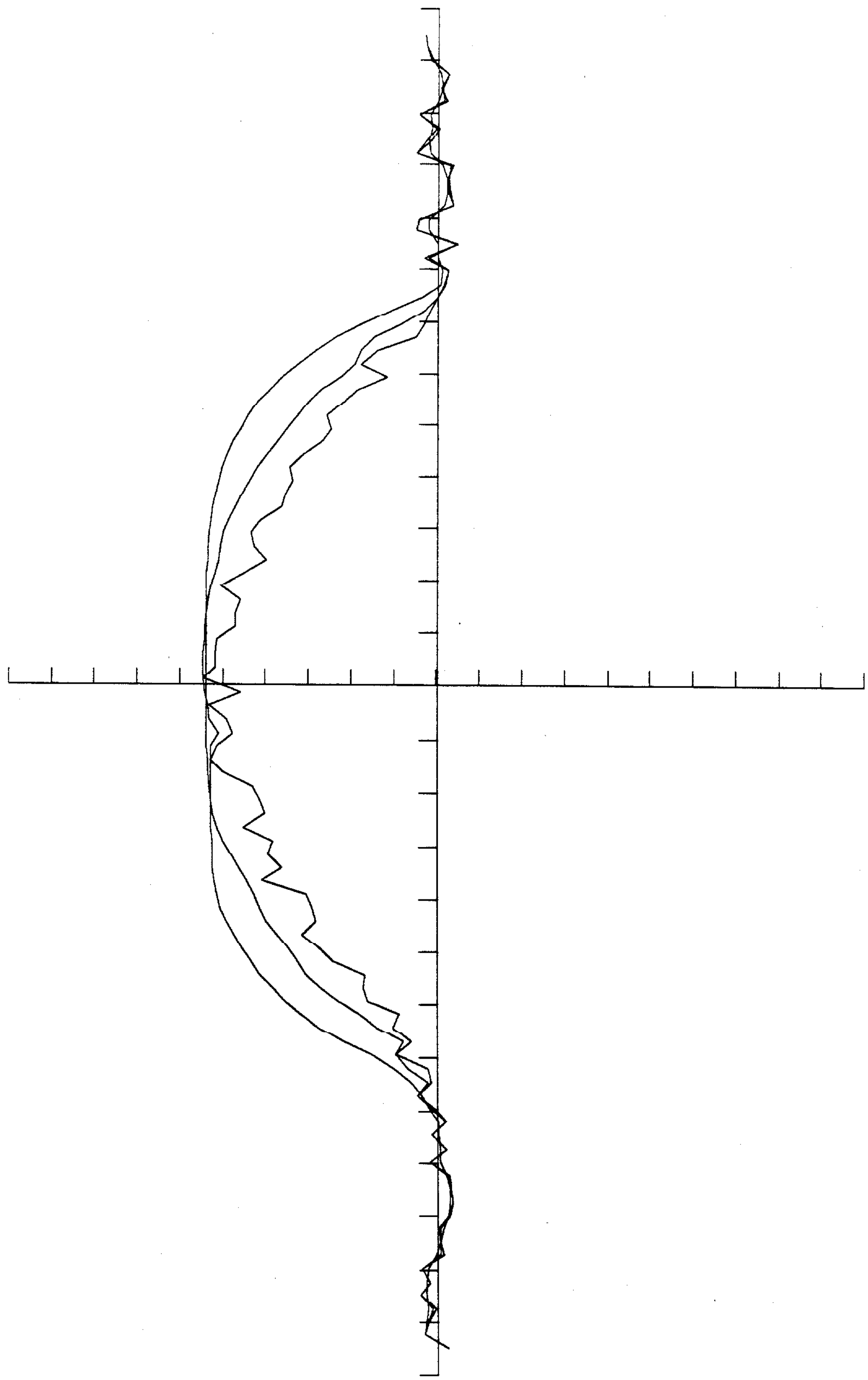
Plot 2



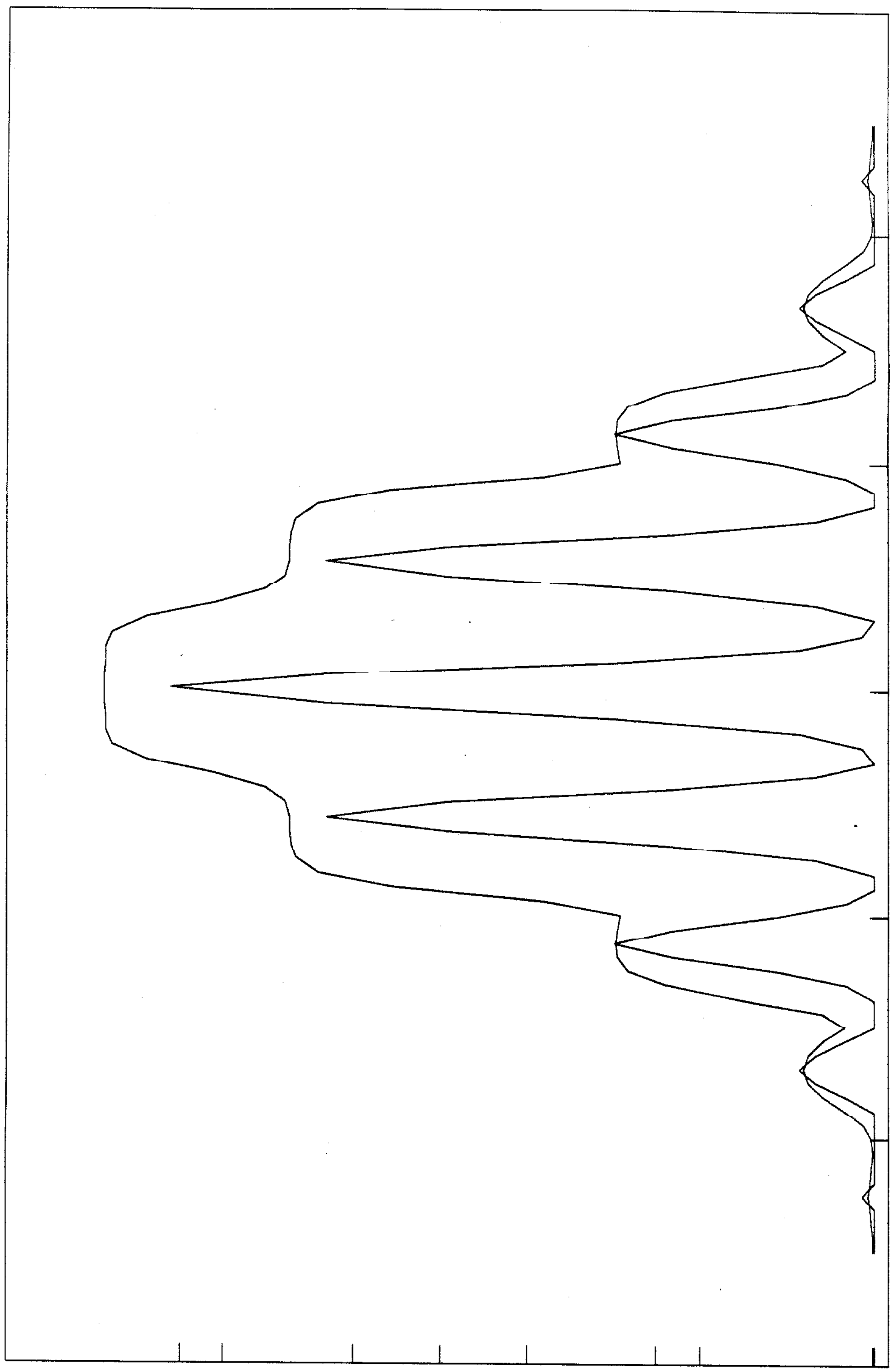
Plot 3



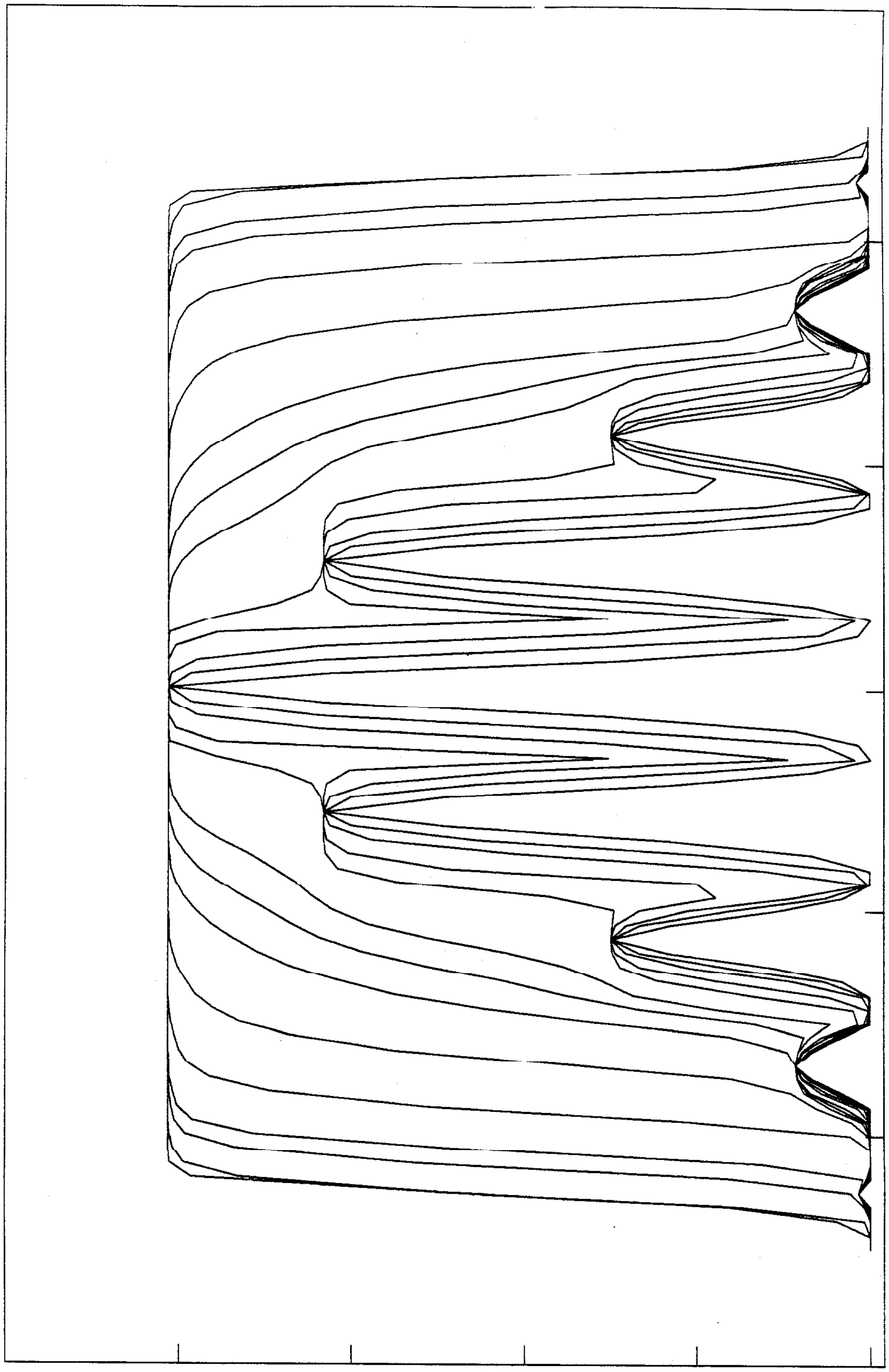
Plot 4



Plot 5



Plot 6



Plot 7

APPENDIX::

```
#define N          100
#include <stdio.h>
#include <math.h>
#include <strings.h>

                /* diffusion constant K = D*dt/dx**2<= 1/2*/
                /* Courant number= dt/dx < 1/max(u) */
double CFL, S;
main(argc,argv)
char **argv;
int argc;
{
FILE    *fp,*fopen();
char    *s = "\0";
int     i,j,I;
double  K, u[N], u1[N],flux(),sign(),F(),aver(),diff(),Flux(),aver(),c
    printf("\n# iterations:  ");
    scanf("%d",&I);
    printf("\n diffusion constant K = ");
    scanf("%lf",&K);
    printf("\n CFL = ");
    scanf("%lf",&CFL);
    printf("\n S = ");
    scanf("%lf",&S);
    fp = fopen(argv[1],"r");
    for(i=0;i<N;++i)
            fscanf(fp,"%lf",&u[i]);
     for(j=0;j<I;++j){
        u1[0] = u[0];
        u1[1] = u[1];
        u1[N-2] = u[N-2];
        u1[N-1] = u[N-1];
        for(i=2;i<N-2;++i)
          u1[i] = u[i]-sign(u[i-1]-u[i+1])
                  * cfl(u,i)*( Flux(u,i+1)-Flux(u,i) )+K*diff(u,i);
        for(i=0;i<N;++i)
          u[i] = u1[i];
      }
      fclose(fp);
      fp = fopen(argv[2],"w");
      for(i=0;i<N;++i)
        fprintf(fp,"%f\n",u[i]);
        fclose(fp);
        strcpy(s,"Plot ");
        strcat(s,argv[2]);
        strcat(s," 100");
        printf("\n   %s\n",s);
        system(s);
}

double  diff(u,i)
double *u;
int i;
{
                return (  (u[i+1] - 2.0*u[i] + u[i-1]) );
}
double sign(x)
double x;
{
        return( x > 0.0 ? 1.0 :(x < 0.0 ? -1.0 : 0.0));
```

APPENDIX:

```
}

double F(u,i)
double *u;
int i;

{
     double aver(), F;

     F = 0.5*u[i]*u[i] - aver(u,i)*u[i] ;
     return(F);
}

double aver(u,i)
double *u;
int i;
{
double fabs();
/*  if ( fabs(u[i] - u[i-1]) > 0.5 )
        return( (u[i]+u[i-1])/2.0 ) ;
 else */
     return( 0.0);

}

double flux(u,i)
double *u;
int i;
{
   double w,f, F(),sign();

  w = sign(u[i-1]-u[i])*( F(u,i) - F(u,i-1) ) * (u[i] - u[i-1]);
   if ( w < 0.0 )
     f = F(u,i);

    else
      f = F(u,i-1);
    return( f );
}
 double Flux(u,i)
double *u;
int i;
{
double flux(), r(), f,sign(),cfl();
/*  if ( r(u,i) > 0.0 && r(u,i-1) < 0.0 ) {
f = 0.5*(flux(u,i+1)+flux(u,i-1) - cfl(u,i)*(u[i] -u[i-1])*(r(u,i)-r(u
  return( f );
  }*/
  f = flux(u,i);
  return( f );

}

double r(u,i)
double *u;
int i;
{
     double aver(),sign(), r;
     r = sign(u[i-1]-u[i])*( u[i] - aver(u,i));
     return( r );
```

APPENDIX:

```
}

double cfl(u,i)

double *u;
int i;
{
double y, fabs(), max();
double extern S;
 y = (-CFL/S)*fabs(u[i+1]-u[i-1]) + CFL;
 return ( max(y,0.0) );
/*return ( fabs(u[i]-u[i-1]) < S ? CFL : 0.0);*/
}

double max(x,y)
double x, y;
{
/*  printf("\n CFL = %f cfl() = %f",CFL,  ( x > y ? x : y )) ;*/
 return ( x > y ? x : y ) ;
}
```

Chapter 2

Surface Distributions and Numerical Analysis of Singularities

> *Form, in the narrow sense, is the boundary between one surface and another: that is its external meaning. But it has also an internal significance, of varying intensity; and properly speaking form is the external expression of inner meaning.*[2]

§2.1. *Some chronology of the subject.*

The purpose of this chapter is to gain an insight into the nature of operations which will eventually lead to the calculation of singularities. There is much more to be gained from this calculation then just the "edge detection," as we will see. In the previous chapter it was ascertained that the singular part of a function is mixed with the regular part. So what we need is a theory which permits us to compute these singular contributions just as if continuous values are being estimated. The whole power of mathematical analysis is available at smooth neighborhoods, with its arsenal of arithmetics, differentiation, integration, series expansions, etc. Only under exceptional circumstances is it prohibited and useless to apply. Alas, it takes place at the precise sites of our inquiry—places of nonsmoothness. Discontinuity of a function is but the simplest kind of nonsmooth conduct we shall pay attention to. Functions may also possess discontinuous derivatives, even hidden beneath the jumps of derivatives of the lower order. The last statement is devoid of sense when the usual notion of derivative is in use. Indeed, if a function has a jump, how can its

[2] Wassily Kandinsky in Concerning the Spiritual in Art.

derivative be defined? Yet Figures 1.2 and 1.3 are just that kind of phenomenon. Clearly, our interest is not in the point value here, but the function's behavior as it approaches the singular site. For this will tell whether it is a jump point, as defined as in Chapter 1. In such a way jump points and their respective unique normal vectors may be computed. This, in view of the synopsis (1.2.3), will amount to detection of the essential boundary of the BV(Ω) function. Observe that the essential boundary refers only to *jump* points of a function, while we also would like to compute singularities of *arbitrary* order! That is, jumps, jumps under the jumps, etc. What is the meaning of all that and what sort of machinery is to be used to solve this novel task? Certainly, not the good old calculus. Is it possible at all? "Biological vision machines" detect singularities without any difficulty. Well, maybe some higher intelligence or even *artificial* one is responsible?

The answer was hinted by the British physicist P.A.M. Dirac in the late 1920's with an introduction of the celebrated Dirac δ-function. It took another 30 years and the genius of L. Shwartz to show that Dirac's intuition was exactly right. He had shown that not only the δ-function is mathematically sound but that the whole calculus is extendible to deal with singularities. Thus, it became possible to differentiate and to integrate discontinuous functions with ease, and sometimes even to multiply and divide. The name is the Theory of Generalized Functions or Distributions Theory (depending on whether you follow the Russian or French schools).

One wonders why it took so long for the "edge detection industry" to discover it? The answer to the mystery is quite simple. A typical view is, "The problem of finding edges is involved because of the fuzziness of the concept 'edge;' in fact, an edge on a digital image is an intuitive concept more than a formal one" [Lev.]. Such an attitude permeates this research field today.

There are, however, a few notable exceptions when researchers have faced the problem. Some tangible experiments have come in [Can.], even though only one-dimensional singularities were treated and extension to 2-d is totally ad hoc.

T. Binford in [Bin.] remarks that "Intensity boundaries are extended two-dimensional step functions, roughly speaking, or delta functions defined along a curve in the continuous plane." However, no use of this profound observation has followed. [Ber.] looked at the accuracy of the so called Laplacian Detectors (which are the crust of [Ma. Hi.] "Theory of edge detection"). The author assumes 2-dimensional discontinuities of the simplest types and arrives at very interesting estimates. Quite a task without using the power of the theory of distributions! And finally [Lec. Zuc.] had understood well the importance of local structure of image discontinuities in one dimension and the inability of Laplacian detectors to deal with it, yet no consistent mathematical model was presented, even for 1-dimensional signals.

Before proceeding further, some statement of the authorship of the results in the following pages is necessary. Except for the introductory paragraph on distributions, all results were obtained by the author unless otherwise stated. Some differentiation formulas of the calculus of distributions were derived earlier in [Est. Kan. 1, 2] via tensor calculations (and applied to wave propagation problems) by a method that differs from mine. My derivations are much more suited to our intended applications.

Simplification is considerable, since the theorem of integration by parts of tangential derivatives is applied in place of a heavy use of tensor calculus.

And now, let us proceed with the brief introduction to the theory of distributions on surfaces.

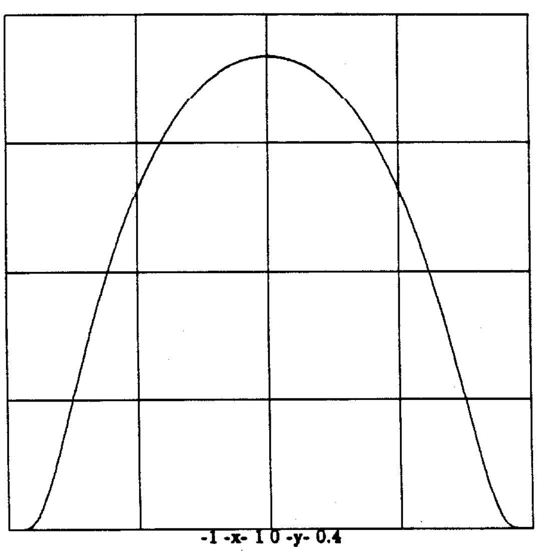
§2.2. *Distributions on surfaces.*

Let $\mathcal{K}_n$ be a space of infinitely differential test functions with compact support on R^n. In particular, if $\varphi \in \mathcal{K}_n$ is identically zero outside some n-dimensional finite region Ω, we say φ is concentrated in Ω. Obviously $\mathcal{K}_n$ is a linear space, though without the usual metrics.

Before giving an exposition of the further notions, let us convince ourselves that $\mathcal{K}_n$ has more elements than just $\varphi \equiv 0$. We write

$$\varphi_N(x) = \begin{cases} \exp\left(-\dfrac{N^2}{N^2 - |x|^2}\right), & \text{for } |x| < N \\ 0, \text{ otherwise} \end{cases} . \tag{2.1.0}$$

Figure 2.1 depicts this test function:



(Fig. 2.1.)

$|x|^2 = \Sigma x_i^2$ and we see that its support is a sphere with radius N. Clearly $\varphi_N(x)$ is an infinitely differentiable function.

We say that a sequence $f_1, f_2, \ldots, f_k, \ldots$ in $\mathcal{K}_n$ converges to 0 in $\mathcal{K}_n$ iff all f_i are identically zero outside a common finite support and $f_i^{(n)}$ converges to zero uniformly for any $n = 0, 1, 2, \ldots$. We also say $\lim f_i = f \in \mathcal{K}_n$ iff $(f_i - f) \to 0$ in the above sense. Every locally integrable function $f(x) : R^n \to R$ gives a rise to the linear continuous functional on $\mathcal{K}_n$,

$$<f, \varphi> = \int_{R^n} f \cdot \varphi \cdot dx_1 \cdot dx_2 \cdot \ldots \cdot dx_n. \tag{2.1.1}$$

$< \ >$ will denote a scalar product of functions.

We shall call $\mathcal{K}_n'$ a space of <u>all</u> continuous linear functionals on $\mathcal{K}_n$, i.e., $g \in \mathcal{K}_n'$, then $g : \mathcal{K}_n \to R$. If g is representable in terms of locally integrable function as in (2.1.1), it is said to be a regular *generalized* function; otherwise it is called a *singular* generalized function. For instance, if we define functional $\delta(x)$ by

$$<\delta(x), \varphi(x)> = \varphi(0), \tag{2.1.2}$$

then $\delta(x)$ is continuous and linear on $\mathcal{K}_1$ and therefore $\delta(x) \in \mathcal{K}_1'$; however it does not correspond to any locally integrable function. Two generalized functions g(x) and f(x) are equal iff $<g, \varphi> = <f, \varphi>$ for any $\varphi \in \mathcal{K}_n$. Evidently, constant functions correspond to constant generalized functions and different *usual* functions correspond to different generalized functions, if they differ on the set of nonzero measure. $\mathcal{K}'_n$ is a linear space on its own, so operations of addition and multiplication by scalar are defined in the obvious manner. The operation of differentiation is what makes the whole construction so useful. Since for usual functions f, φ

$$\left\langle \frac{\partial f}{\partial x_i}, \varphi \right\rangle = \int_{R^n} \frac{\partial f}{\partial x_i} \varphi \, d\bar{x} , \tag{2.2}$$

using integration by parts

$$= \int_{R^{n-1}} \left\{ f \cdot \varphi \Big|_{x_i=-\infty}^{+\infty} f\varphi \right\} d\bar{x} - \int_{R^n} \frac{\partial \varphi}{\partial x_i} \cdot f d\bar{x} = -\left\langle f, \frac{\partial \varphi}{\partial x_i} \right\rangle ,$$

since φ has bounded support and vanishes on $\pm\infty$. By (2.2), differentiation of the regular function $f(\bar{x})$ is transformed into the differentiation of the test function. Clearly, if $f(\bar{x})$ were not differentiable at all, still we could define $\left\langle f, \frac{\partial \varphi}{\partial x_i} \right\rangle$. The usefulness of such *formal* differentiation is demonstrated by the examples below.

Let

$$h(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases} . \tag{2.3}$$

Then by (2.2)

$$\langle h'(x), \varphi(x) \rangle = -\langle h(x), \varphi'(x) \rangle .$$

We have from (2.1.1)

$$= -\int_{-\infty}^{+\infty} \varphi'(x) \cdot h(x) \, dx = -\int_{0}^{\infty} \varphi'(x) \, dx = \varphi(0), \tag{2.4.1}$$

because $\varphi(\infty) = 0$, $\varphi(x)$ being of finite support.

In view of (2.1.2) it is true that for any $\varphi \in \mathcal{K}_1$

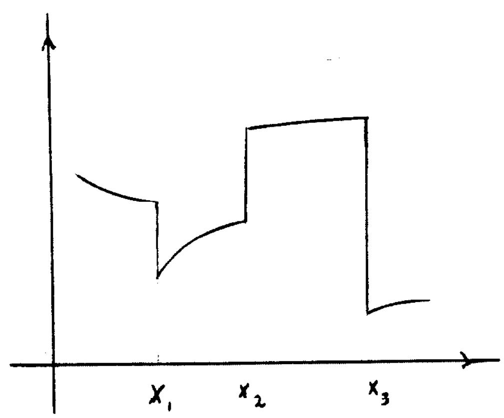
$$<h'(x), \varphi(x)> \;=\; <\delta(x), \varphi(x)>,$$

which yields

$$h'(x) = \delta(x) \tag{2.4.2}$$

upon application of the equality condition for generalized functions.

In the next example, Lebesgue's decomposition of the derivative of a function of a single variable is computed. Let f(x) be an arbitrary discontinuous function, whose graph is on Figure 2.2.



(Fig. 2.2)

Suppose that f'(x) is piecewise differentiable and x_1, x_2, x_3, ..., x_m are its discontinuity points. Now, let us extract singularities from f(x) as follows:

$$g(x) = f(x) - \sum_m H_m \cdot h(x - x_m), \tag{2.5}$$

where $H_m = f^+(x_m) - f^-(x_m)$, i.e., a jump in x_m. Equation (2.5) shows that $g(x)$ is a continuous differentiable function and is the "cut and paste" version of the discontinuous function $f(x)$. It suffices to differentiate $g(x)$ in (2.5) term by term and use (2.4.2) to obtain

$$g'(x) = f'(x) - \sum_m H_m \cdot \delta(x - x_m), \tag{2.6}$$

where $f'(x)$ is a *generalized* derivative of $f(x)$ in the sense of (2.2). We establish that

$$f'(x) = g'(x) + \sum_m H_m \cdot \delta(x - x_m). \tag{2.7}$$

Thus Lebesgue's decomposition of the $f'(x)$ is concluded as claimed, $g'(x)$ representing absolutely continuous contribution and $\sum_m H_m \cdot \delta(x - x_m)$ corresponding to the pure point contribution.

Before going into the subject of distributions concentrated on surfaces, some discussion of local properties of generalized functions is worthwhile, for it is relevant to the subject of detecting nonregular parts of a generalized function. A generalized function is a functional, i.e., it takes a function and makes a number. So strictly speaking it has no behavior at the point, nor it has point value $g(x_0)$. However, the notion of *locality* could extend to the functional via the test function on which it acts. If it happens that for some point x_0 and any test function φ_{x_0} whose support is only in the neighborhood of x_0, $\langle g, \varphi_{x_0}\rangle = 0$, we say that $g(x)$ is zero in the vicinity of x_0. We may say that the generalized function g(x) is equal to zero in the region Ω if for any $\varphi \in \mathcal{K}^n$ such that $\mathrm{supp}(\varphi) \subset \Omega$, it follows that $<g, \varphi> = 0$, where $\mathrm{supp}(\varphi)$ denotes the set $\{x \mid \varphi(x) \neq 0\}$. This is equivalent to stating that g(x) is zero in the neighborhood of each point *inside* region Ω.

If a functional g(x) is equal to zero everywhere in R^n except for some region Ω, i.e., $\mathrm{supp}(g) = \Omega$, then g(x) is said to be *concentrated* on the set Ω. For example, a regular functional f(x) which corresponds to the usual continuous function f(x) may be nonzero at some x_0 as a generalized function, while $f(x_0) = 0$ in the usual sense. Another important example is furnished by the $\delta(x)$ function, for $\mathrm{supp}(\delta(x)) = \{0\}$, i.e., a set consisting of just one origin point. In such a way two generalized functions can be compared for equality, if their difference is zero in R^n in the above connotation. In sum, generalized functions which are equal in every neighborhood are equal. Moreover, all of their *generalized* derivatives are also equal. Now we are ready to state the main result on local structure of generalized functions.

Theorem 2.1 (Gelfand) [Gel. Shil.].

For any generalized function g(x), s.t. $\mathrm{supp}(g(x)) \subset \Omega \subset R^n$, there exists such number m and regular functions $g_k(x)$ that for any test function $\varphi(x)$, concentrated in Ω, it follows,

$$<g(x), \varphi(x)> = \sum_{|k| \leq m} <g_k(x), \mathcal{D}^k\varphi(x)> , \tag{2.8}$$

i.e., $g(x) = \sum_{|k| \leq m} \mathcal{D}^k g_k$ for some regular functions $g_k(x)$. Here,

$$\mathfrak{D}^k = \frac{\partial^{|k|}}{\partial x_1^{k_1}\, \partial x_2^{k_2} \ldots \partial x_n^{k_n}}, \quad |k| = \sum_{i=1}^{n} k_i .$$

This means that any generalized function concentrated on the closed region Ω is just a linear combination of derivatives of some continuous functions, which vanish outside arbitrary ϵ-neighborhood of Ω. No matter how complicated the nature of a generalized function is, it is simply a sum of the derivatives of usual functions. This allows us to have global classification of the whole space $\mathcal{K}_n'$ given by the following definition:

Definition 2.1.

$g \in \mathcal{K}_n'$ has order of singularity m if it can be represented as

$$<g(x), \varphi(x)> = \sum_{|k| \leq m} \int_{R^n} g_k \cdot (x) \cdot \mathfrak{D}^k \varphi(x) dx , \qquad (2.9)$$

where all $g_k(x)$ are usual functions on R_n and m is the smallest such number.

Definition 2.1 and Theorem 2.1 imply the following corollary:

Corollary 2.1.

Every generalized function with compact support is of *finite* singularity order.

We next introduce a concept of the generalized function concentrated on the surface in R^n.

<u>Definition 2.2.</u>

Generalized function g is said to be *concentrated on the surface* L if

$$<g, \varphi> = \int_L g(x) \cdot \varphi(x) \, d\ell , \tag{2.10}$$

where $\varphi(x) \in \mathcal{K}_n$ is any test function and $d\ell$ is a surface area element of the surface L in R^n. The dimension of the surface L is less or equal to $(n - 1)$. This definition could be generalized to the higher orders of singularity by

$$<g, \varphi> = \int_L \sum_{i=0}^{n} g_i(x) \cdot \mathfrak{D}^i\varphi(x) \, d\ell , \tag{2.11}$$

which is of n-th singularity order.

Surface distributions (2.10), (2.11) are the most important concepts in our study of decomposing functions into the sum of regular and the singular parts. It can be shown [Gel. Shil.] that any generalized function concentrated in the point is a linear combination of δ-function and its derivatives $\delta^{(n)}$, $n = 1, 2, \ldots$, see Theorem 2.1 above. Derivatives of $\delta(x)$ are defined by

$$<\delta^{(n)}(x), \varphi(x)> = (-1)^n < \delta(x), \varphi^{(n)}(x)> . \tag{2.12}$$

Hence, by the definition of $\delta(x)$, (2.12) can be rewritten in the form

$$= (-1)^n \varphi^{(n)}(0),$$

i.e., n-th derivative of the $\delta(x)$ is a *functional* which estimates the n-th derivative of the test function at the origin.

Similarly, we have for shifted $\delta^{(n)}$-function:

$$<\delta^{(n)}(x - x_0), \varphi(x)> = (-1)^n \varphi^{(n)}(x_0) \ . \tag{2.13}$$

For distributions of several variables concentrated on surfaces, an equivalent role is played by the so called *layers*. The *simple layer* $g(x) \cdot \delta_L$ is given by its action on the test functions space $\mathcal{K}_n$ by (2.10), where g(x) is called a *density* of the simple layer over the surface L. The *double layer* involves differentiation of the test function in the direction normal to the surface L, so we have

$$\left\langle \frac{\partial}{\partial n} f(x) \cdot \delta_L, \varphi(x) \right\rangle = - \int_L f(x) \cdot \frac{\partial}{\partial n} \varphi(x) \cdot d\ell. \tag{2.14}$$

and f(x) is its density.

More generally, *k-th order multi-layer* with the density f(x) is produced by the equation

$$\left\langle \frac{\partial^k}{\partial n^k} [f(x) \cdot \delta_L], \varphi(x) \right\rangle = (-1)^k \int_{\substack{L \\ k=0,1,2,\dots}} f(x) \cdot \frac{\partial^k}{\partial n^k} \varphi(x) \cdot d\ell \ . \tag{2.15}$$

Parallel with the formula, (2.13) is selfevident. By $\frac{\partial^k}{\partial n^k}$ we understand k-times differentiation in the direction *normal* to the surface in *each* point of the integration (2.15). Examination of (2.15) shows that the definition of the k-th layer is independent from the coordinate system in which surface L is expressed in R^n. The alternative approach to the generalized functions on surfaces was developed by I. Gelfand [Gel. Shil.], in which equation $P(\overline{x}) = 0$ of the surface L is involved explicitly. It is worthwhile to give at least some definitions of Gelfand's construction, for it may provide a somewhat different approach than taken in this work.

Definition 2.3.

Let $P(\overline{x}) = 0$ be an equation of the surface L in R^n, where $\overline{x} = (x_1, x_2, \ldots, x_n)$. Heaviside function H(x) is given by

$$H(x) = \begin{cases} 1, & \text{for all } i,\ x_i \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$<H(P), \varphi(x)> = \int_{P(x) \geq 0} \varphi(\overline{x})\, d\overline{x}.$$

Then delta function on the surface $P(\overline{x}) = 0$ is set out by

$$\delta(P) = \lim_{c \to 0} \frac{1}{c} [(H(P + c) - H(P)] . \tag{2.16}$$

The same is true with its derivatives

$$\delta^{(k)}(P) = \lim_{c \to 0} \frac{1}{c} [\delta^{(k-1)}(P + c) - \delta^{(k-1)}(P)] . \tag{2.17}$$

Functionals $\delta^{(k)}(P)$ then play the same role in the representation of the singular distribution on the surface P as derivatives $\delta^{(k)}(x)$ in the one-dimensional case of distribution concentrated in the point. However, to see the action of $\delta^{(k)}(P)$ on the arbitrary test function, we need to introduce a local coordinate system whose coordinate unit vectors lie in the plane tangent to the surface $P(x) = 0$ for each x. This will transform (2.17) into the following formula:

$$\langle\delta^{(k)}(P), \varphi(x)\rangle = (-1)^k \int_{P(x)=0} [\partial^{(k)}\left(\varphi \left| \frac{\partial x}{\partial u} \right|\right) / \partial u_1^k]_{u_1=0}\, du_2 du_3 \ldots du_n\,, \qquad (2.18)$$

where local coordinates are chosen by $u_1 = P$, and $u_i = x_i$ for $i = 2, 3, \ldots, n$. It is always possible to find such a representation if the surface is nonsingular, i.e., $|\mathbf{grad}\ P| \neq 0$ on $P(x) = 0$, since the normal vector is given by $\bar{n} = \nabla P / |\nabla P|$ and each x_i could be thus expressed as a function of $\bar{u} = (u_1, u_2, \ldots, u_n)$. Comparing (2.18) and (2.15) we observe then that the k-th derivative of the δ-function on the surface $P(\bar{x}) = 0$, i.e., $\delta^{(k)}(P)$, coarsely resemblesthe k-th multilayer on the same surface $\frac{\partial^k}{\partial \mathbf{n}^k}\, \delta_L$, with the presence of the local coordinates $\{u_i\}$ and the Jacobian $\left|\frac{\partial x}{\partial u}\right|$ making some notable difference. Clearly, change of the original coordinates $(x_1, x_2, \ldots, x_n)$ should change the representation of the surface $P(\bar{x}) = 0$; i.e., expressions in the equation (2.18) appear to be dependent on the particular coordinate system (e.g., $\left|\frac{\partial x}{\partial u}\right|$ might change). Thus it looks as if $\delta^{(k)}(P)$ is not uniquely determined by (2.18). Gelfand shows that this is not so [Gel. Shil] (Vol. 1, Chapter 3), by proving that there exists a unique differential form w of the $(n - 1)$ order invariantly related to the surface $P(\bar{x}) = 0$. Gelfand's $\delta^{(k)}(P)$ have very nice analytical properties, most significantly the following chain rule:

$$\frac{\partial}{\partial x_i}\, \delta^{(k)}(P) = \frac{\partial P}{\partial x_i} \cdot \delta^{(k+1)}(P)\,. \qquad (2.19)$$

Yet, the action of $\delta^{(k)}(P)$ on the test function is not as obvious as the action of the multilayer (2.15). On the other hand this action will guide the analysis of the singularity detector and therefore its construction.

While some useful properties of surface distributions have been established, they remain the most difficult and least developed part of the theory of generalized functions. In particular, very little is known about how to *multiply* surface distributions or how to deal with the distributions defined on the *surfaces with singularities*, i.e., $|\nabla P(\overline{x})| = 0$ for some points of $P(\overline{x}) = 0$. These limitations put severe obstacles into the author's inquiry about singularity detection, as we shall see. Specifically, analysis of the "interesting" points on the edge, such as corners, curvature brakes, etc., demands new theoretical development for the surfaces with singularities. Similarly it will become of practical importance here to be able to multiply multilayers by multilayers in order to analyze nonlinear detection procedures such as the Beaudet corner detector (see [Bea.]) and others.

§2.3. *Fundamental theorem of edge detection.*

Practically every stage of image analysis involves edge extracting as an integral part of the image processing. For example, image segmentation attempts to subdivide an image into uniform connected subregions and then identify the subregions with a semantic description, e.g., a type of aircraft, a geological formulation, a human face, etc. But to find the region, its boundary must be located first by means of edge detection. The shapes of edges provide crucial clues about the shapes of objects. The applications of edge detection could take hundreds of pages to list. As was concluded earlier in Chapter 1, such a multitude of applications results from the fact that singularities and their metamorphoses furnish computational clues about *intrinsic* properties of the geometry of the visual world around us. A great variety of ad hoc edge extraction techniques are used and new ones are developed. One is almost lost in trying to see a common thread in this mixed 'bag of tricks.' Most of these detection algorithms can be expressed in terms of digital integration with *sliding* masks of various types (the term 'convolution' is usually used, which is not exactly proper, since the mask need not possess any symmetries). Here are some examples from earlier and still widely used edge detectors.

Integrating masks are given in their matrix form with the (i, j) indices corresponding to the x and y coordinates respectively.

Roberts detector is given by two masks:

$$R_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; \quad R_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Sobel edge detector:

$$S_{vertical} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}; \quad S_{horizontal} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

Subscripts indicate the direction of maximal response.

Here are some of the rotations of the so-called compass gradient detector:

$$C_{east} = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix}; \; C_{south} = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

There are altogether eight masks—for North, Northeast, East, Southeast, South, Southwest, West and Northwest.

To detect more complicated features, such as lines, *Laplacian* masks are applied, which are digitized finite difference approximations to the Laplacian operator:

$$\text{Laplacian} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}.$$

And so on and on ... . Some authors derive their masks from the explicit use of finite differences of various types, some carry out differences on the averaged, i.e., smoothed out, images and others describe their masks as means of matching, correlating the detector's layout with the structure of the edge. These last are usually classified as detection through fitting [Na. Bi.]. In yet another example, the entire "theory" of computational vision has been developed in a number of leading institutions (e.g., MIT, CMU), based on the notion of zero-crossings, i.e., locating output zeros of convolutions with miscellaneous Laplacian-like masks [Ma. Hi.].

An essential first step, as a preliminary to the discussion of the Fundamental Theorem of Edge Detection, is to observe that the design of the detecting schemes above originates from the vague premise of computing in some measure the *size of the jump* of the prospective edge at the site where the mask is centered. This is thought to be especially true when the edge is separating two regions of uniform brightness. A literature review will reveal that *no formal proof* that the detector actually computes some attribute of the *edge* has ever been given, with the exception of [Ber.] where only trivial discontinuity models were accounted for by the Laplacian zero-crossing detector.

It hardly needs to be said that this is an intolerable situation and has to be dealt with to accomplish any progress in the pursuit of establishing the foundations for the "*Numerical analysis of singularities.*"

Fortunately the calculus of generalized functions provides direct techniques for computing singularity's contribution to the "melting pot" of integration with a smooth function. In effect, we shall try to separate the contribution of the *edge* from that of smooth gradients in the image, called *low frequency* noise [Pavl.]. Essentially, we have already demonstrated such a technique by decomposing the arbitrary piecewise differentiable function of one variable in the sum of continuous and step functions in (2.5), or equivalently separating its *generalized* derivative into continuous part plus linear combination of shifted δ-functions in (2.7). If $f'(x)$ in (2.7) is to be integrated with any smooth function $\varphi(x)$, then (2.7) becomes

$$<f'(x),\varphi(x)> = \int_{-\infty}^{+\infty} g'(x)\cdot\varphi(x)\cdot dx + \int_{-\infty}^{+\infty}\Big[\sum_m H_m\cdot\delta(x-x_m)\Big]\cdot\varphi(x)\cdot dx\ , \quad (2.20)$$

by (2.4.1)

$$= <g'(x),\ \varphi(x)> + \sum_m H_m\cdot\varphi(x_m)\ .$$

And so we see that $\sum_m H_m \cdot \varphi(x_m)$ is the so much desired contribution of the *singular* points x_m.

Accordingly, the underlying principle of identification of these discontinuity points is in designing a set of masks $\varphi_i(x)$ in such a way that the regular part $<g', \varphi>$ of $<f', \varphi>$ in (2.20) is negligible compared to the singular term. The grand strategy then is to annihilate smooth contribution and expose singularities when they react, i.e., precipitate nonzero contribution into the singular term.

This is easily achieved by the *convolution* of $f'(x)$ with a simple test function $\varphi(x)$ defined in (2.1.0), where $\epsilon > 0$ is taken to be a very small number. Then

$$<g', \varphi> = \int_{-\epsilon}^{+\epsilon} g' \cdot \varphi \cdot dx = O(\epsilon), \tag{2.21}$$

i.e., it is very small due to the fact that $g'(x)$ and $\varphi(x)$ are both bounded functions and the interval of integration is shrinking to zero. On the other hand the singular contribution $H_m \cdot \varphi(x - x_m)$ is independent of the size of the supp($\varphi(x)$), i.e., it does not change *at all* when $[-\epsilon, +\epsilon] \rightarrow (\{0\})$ and is only subject to the size of the jump of the function $f(x)$ at the point x_m, namely H_m. Obviously, it is a desired dependency when one is only interested in finding jumps *and their values*. Henceforth functions of the type $\varphi_\epsilon(x)$ shall be called *thin masks*, for they are built on a very narrow support. The above strategy presumably solves the problem of locating simple discontinuities of the arbitrary piecewise smooth function of one variable $f(x)$. So, are we ready to write the computer program to implement such an algorithm? A little reflection on the meaning of (2.20) yields a negative answer! Observe that (2.20) decomposition deals with the $f'(x)$, i.e., the derivative of $f(x)$ which does not exist in the usual sense. In fact, $f'(x)$ is a *generalized* derivative of $f(x)$. It is therefore a *functional* on the space $\mathcal{K}_1$ and generally has no point values which could be inputted into the computer memory. Consequently, $<f'(x), \varphi(x)>$ of (2.20) appears to be noncomputable, rendering all considerations above worthless.

By good fortune, the foregoing construction is rescued via transferring differentiation of $f(x)$ in (2.20) to the differentiating of the test function $\varphi(x)$, since according to (2.2) the following identity holds even for *generalized* derivatives:

$$<f'(x), \varphi_\epsilon(x)> \; = -<f(x), \varphi'_\epsilon(x)> . \tag{2.22}$$

It is clear now that the computer algorithm can be designed along the lines suggested previously in the discussion of (2.20) by means of the *thin mask* φ_ϵ. Incidentally, in contrast with $f'(x)$, $\varphi'_\epsilon(x)$ is an ordinary function, whose point values are easily digested by the computer.

Ignoring such "minor matters" as the presence of noise, digital sampling and what it can do to the signal, we just have resolved the problem of detecting discontinuities in one-dimensional waveforms, guided by the exact knowledge of the contribution made by point singularity to the local integration with a smooth mask.

But how about singularities of functions of several variables—is it possible to establish a decomposition formula which is a generalization of (2.20)? If the answer is positive, then we will have a *constructive* understanding of how to interact with singularities, which will lead the way to the computational analysis of existing edge detectors as well as the design and analysis of totally new algorithms.

The starting point is to define the "smoothest" kind of discontinuous functions of several variables. It is formulated in line with Kolmogorov's notion of a set of points with normal jump (section 1.2.2) as well as Volpert's *essential boundary* in Theorem 1.5.

Definition 2.4.

Function $f(x_1, x_2, \ldots, x_n) \in BV(\Omega)$ is said to be a *smooth function with an essential singularity,* if there exists a set $\Gamma(f)$, which is coverable by a finite number of essential boundaries of sets with finite perimeters (see Definition 1.6), such that $f(x_1, x_2, \ldots, x_n)$ is C_∞ everywhere except for the set $\Gamma(f)$, which comprises its *essential* boundary or *set of jumps*. If $\Gamma(f)$ forms a surface L in Ω, then we write $[f]_L = |\Delta| f_L$ for the jump defined in (1.9.1) of the $f(\overline{x})$ across L in the direction of outward normal, as in the discussion of [Vol.] in (1.2.2).

This lengthy definition is a fancy way to say that $f(\overline{x})$ is piecewise smooth except for some *surface* $\Gamma(f)$, across which it has a jump discontinuity. However, the rigorous definition is required to be able to consider very complicated and nonsmooth boundaries, still treatable as long as they consist of *jump points*, modulo set of "length" zero, with their respective unique *normals* in the sense defined previously in Chapter 1. And $[f]_L$ is a jump function defined only on the boundary L and is equal at each point $x_0 \in L$ to the difference between the *outside* and the *inside* one-sided limits of the function $f(\overline{x})$ at x_0 on the surface L.

In order to avoid possible confusion, we shall distinguish between a *generalized* derivative and its regular component.

Definition 2.5.

For any $f(\overline{x}) \in \mathcal{K}'_n$, symbol $\frac{\overline{\partial f(\overline{x})}}{\partial x_i}$ will denote a *generalized* derivative of $f(\overline{x})$ as in (2.2). And symbol $\frac{\partial f(\overline{x})}{\partial x_i}$ shall be the *regular part* of $\frac{\overline{\partial f(\overline{x})}}{\partial x_i}$ as in (2.20).

The ramifications of the following simple theorem to the Numerical Analysis of Singularities are of such magnitude that I have decided to call it here the Fundamental Theorem of Edge Detection or FTED. The reader should judge if it is so.

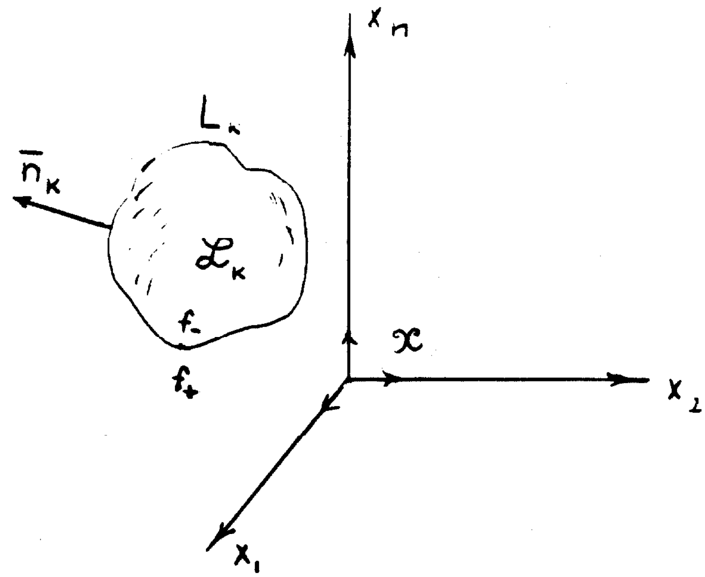
The proof is based on the application of Green's theorem to generalized partial differentiation, which is routinely used in any text on the Theory of Generalized Functions. Therefore, I shall not attribute it to any author.

Theorem 2.2 (Fundamental Theorem of Edge Detection).

Let $f(\overline{x}) \in BV(\Omega)$ be a smooth function with essential singularity comprised by the $\bigcup_K L_K$ of essential boundaries $L_K \subset \Omega \subset R^n$. And let n_K be an outward unit normal to L_K and $\{x_i\}$ be an orthnormal basis of R^n. Then we may write

$$\frac{\overline{\partial f}}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_K n_K \bullet x_i \cdot [f]_{L_K} \cdot \delta_{L_K} . \tag{2.23}$$

Proof:



(Fig. 2.3)

Without loss of generality let us depict the surface boundary in Figure 2.3 by the closed surface L_K. If the essential boundary has a boundary of its own, the derivation will be still valid since $[f]_{L_K}$ will simply vanish on the complement of L_K to any boundaryless surface.

Let $\varphi(x_1, x_2, \ldots, x_n) \in \mathcal{K}_n$ be an arbitrary test function. Then by (2.2)

$$\left\langle \overline{\frac{\partial f}{\partial x_i}}, \varphi \right\rangle = -\left\langle f, \frac{\partial \varphi}{\partial x_i} \right\rangle = -\int_{\Omega} f \cdot \frac{\partial \varphi}{\partial x_i} \cdot dx_1 \cdot dx_2 \cdot \ldots \cdot dx_n$$

$$= -\left[\int_{\Omega/\mathcal{L}} f \cdot \frac{\partial \varphi}{\partial x_i} \, d\overline{x} + \int_{\mathcal{L}} f \cdot \frac{\partial \varphi}{\partial x_i} \, d\overline{x} \right],$$

by the application of Green's theorem (1.2.8)

$$= \int_{\Omega \setminus \mathcal{L}} \frac{\partial f}{\partial x_i} \cdot \varphi \, d\overline{x} + \int_{L_K} f^+ \cdot \varphi \cdot \cos(n_k \hat{\ } x_i) \, d\ell \, +$$

$$+ \int_{\mathcal{L}} \frac{\partial f}{\partial x_i} \cdot \varphi \cdot d\overline{x} - \int_{L_K} \overline{f} \cdot \varphi \cdot \cos(n_k \hat{\ } x_i) \, d\ell \, ,$$

then collecting similar terms we get

$$= \int_{\Omega} \frac{\partial f}{\partial x_i} \cdot \varphi \cdot dx_1 \cdot dx_2 \cdot \ \ldots \cdot dx_n + \int_{L_K} (f^+ - f^-) \cdot \varphi \cdot n_k \bullet x_i \, d\ell \,,$$

where $d\ell$ is L's surface differential, and with the (2.10) definition of a simple layer it is rewritten

$$= \left\langle \frac{\partial f}{\partial x_i} , \varphi \right\rangle + <[f]_{L_K} \cdot n_k \bullet x_i \cdot \delta_{L_K}, \varphi> .$$

Next we sum inner products,

$$= \left\langle \frac{\partial f}{\partial x_i} + n_k \bullet x_i \cdot [f]_{L_k} \cdot \delta_{L_K}, \varphi \right\rangle .$$

Hence, using the definition of equality of generalized functions and observing that $\varphi(\overline{x})$ was an arbitrary test function, we conclude that

$$\frac{\overline{\partial f}}{\partial x_i} f = \frac{\partial f}{\partial x_i} + n_k \bullet x_i \, [f]_{L_K} \cdot \delta_{L_K}$$

which yields the statement of Theorem (2.23) upon induction argument on the number of disjoint boundaries L_K's. Q.E.D.

<u>**Remark:**</u>

Identity (2.23) states the equality of *functionals*, even though $\frac{\partial f}{\partial x_i}$ is just a usual function, i.e., a *regular* functional.

The fundamental point of Theorem 2.2 is in asserting that essential singularities of the discontinuous function of several variables act as *simple layers* concentrated on the essential boundary, with the density equal to the size of the jump across the discontinuity. Theorem 2.2 is a Fundamental Theorem of Edge Detection because it fully describes the contribution made by the edge-like singularity of the most general nature into the inner product with an arbitrary smooth mask.

The consequences of (2.23) are so far-reaching that not only we will be able, for the first time, to make intelligent statements about existing edge, line, corner, zero-crossings, motion, etc. detectors, but also totally new concepts will flow out of its application. In fact, it will become clear that this new branch of Numerical Analysis, which I call Numerical Analysis of Singularities, awaits to be unfolded. Its methods will be of a mathematical nature and not the astrology of "edge detection industry" of the past 30 years.

§2.4. *Direction of the generalized gradient in the neighborhood of singularity.*

Numerous popular edge detection schemes explicitly use the assumption that the direction of the gradient of the image (or smoothed image) is normal to the direction of the potential edge. For example, in [Har.] "a pixel is marked as an edge pixel if in the pixel's immediate area there is a zero-crossing of the second directional derivative in the direction of the gradient ..." We shall later return to the "second directional derivative" part. Our specific purpose here is to exploit the "direction of the gradient" idea. A similar viewpoint on the direction of the gradient being normal to the edge's direction is employed in the so-called Nonmaximum Suppression scheme in [Can.] with distinctly good experimental results.

To analyze the local behavior of the direction of the gradient near and at the essential discontinuity point, we employ Theorem 2.2, FTED. Note that the gradient does not exist in the regular sense at the discontinuity sites, so we shall use its *generalized* counterpart. For the moment we consider only smooth functions $f(\overline{x})$ with essential singularity, as in Definition 2.4. When such a function is integrated with the sliding mask $\varphi(\overline{x})$, the result is expressible in terms of the *convolution* $f(\overline{x}) * \varphi(\overline{x})$, where

$$f(\overline{x}) * \varphi(\overline{x}) = \int_{\Omega} f(\overline{y}) \cdot \varphi(\overline{x} - \overline{y})\, d\overline{y} \tag{2.24}$$

and $\overline{x}$, $\overline{y}$ are n-dimensional vector arguments.

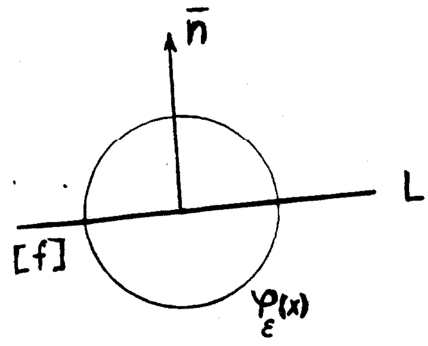
<u>Lemma 2.2.1</u>.

Suppose $f(x_1, x_2, \ldots, x_n)$ to be a smooth function with essential singularity L, of zero curvature, and let $\varphi(x_1, x_2, \ldots, x_n) \in \mathcal{K}_n$. If $\mathrm{supp}(\varphi(\overline{x}))$ is small or if the *regular* part (see Definition 2.5) of the generalized gradient

$$\nabla_{reg} f(\overline{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

is negligent, then the direction of the gradient $\nabla(f(\overline{x}) * \varphi(\overline{x}))\,(\overline{x}_0)$ at $x_0 \in L$ is *normal* to the essential boundary L.

Proof:



(Fig. 2.4)

Without loss of generality, assume that $\varphi(\bar{x})$ is a familiar cup-shaped $\varphi_\varepsilon(\bar{x})$ of (2.1.0) as in Figure 2.4. The k-th component of the gradient vector at $\bar{x}_0 \in L$ is equal to

$$\frac{\partial}{\partial x_k} (f(\bar{x}) * \varphi(\bar{x}))\, (\bar{x}_0) \, . \tag{2.25}$$

Using the formula of differentiation of the convolution, we obtain

$$= \left[\frac{\overline{\partial f(\bar{x})}}{\partial x_k} * \varphi(\bar{x}) \right] (\bar{x}_0) \, .$$

Then using the formula (2.23) of FTED we obtain the following

$$= \int_\Omega \frac{\partial f(\bar{x})}{\partial x_k} \cdot \tilde{\varphi}(\bar{x})\, d\bar{x} + \int_L \cos(\bar{n}\; x_k) \cdot [f] \cdot \tilde{\varphi}(\bar{x}) \cdot d\ell \, . \tag{2.25.1}$$

Where to simplify the notation $\tilde{\varphi}(\bar{x})$ is introduced and is equal to $\varphi(\bar{x} - \bar{x}_0)$, i.e., $\varphi(\bar{x})$ shifted to the point $\bar{x}_0$ on top of the essential boundary L. The hypothesis of the Lemma implies that either supp($\tilde{\varphi}$(x)) or max $\left|\frac{\partial f(\overline{x})}{\partial x_k}\right|$ is small; hence

$$\int_{\Omega} \frac{\partial f(\overline{x})}{\partial x_k} \cdot \tilde{\varphi}\ (\overline{x})\ d\overline{x} = \int_{\text{supp}(\tilde{\varphi}(\overline{x}))-B_\epsilon} \frac{\partial f(\overline{x})}{\partial x_k} \cdot \tilde{\varphi}(\overline{x})\ dx$$

$$\leq \text{Max} \left|\frac{\partial f(\overline{x})}{\partial x_k}\right| \cdot \int_{B_\epsilon} \tilde{\varphi}\ (\overline{x})\ d\overline{x} \approx 0. \tag{2.26}$$

Then (2.25) and (2.26) lead to

$$\frac{\partial}{\partial x_k} (f * g)\ (x_0) = \int_L \cos(\overline{n}\,\hat{}\, x_k) \cdot [f] \cdot \tilde{\varphi}(\overline{x})\ d\ell.$$

Since L has zero curvature, cos ($\overline{n}$ x_k) = const along the portion of the boundary L within supp($\tilde{\varphi}$) = B_ϵ. So we write

$$= \cos(\overline{n}\,\hat{}\, x_k) \cdot \int_L [f] \cdot \tilde{\varphi}(\overline{x})\ d\ell\ , \tag{2.27}$$

which yields an expression for the gradient vector at $\overline{x}_0$, upon factoring out the common constant $\int_L [f] \cdot \tilde{\varphi}(\overline{x})\ d\ell$ and writing

$$\textbf{grad}(f * \varphi)\ (x_0) = \left(\int_L [f] \cdot \tilde{\varphi}\ (\overline{x})\ d\ell\right) \left(\cos\ (\overline{n}\,\hat{}\, x_1), \cos(\overline{n}\,\hat{}\, x_2), \ldots, \cos(\overline{n}\,\hat{}\, x_n)\right). \tag{2.28}$$

But ($\overline{n} \bullet \overline{x}_1$, $\overline{n} \bullet \overline{x}_2$, ..., $\overline{n} \bullet \overline{x}_n$) *is* a unit vector normal to the boundary L at $\overline{x}_0$, where $\{\overline{x}_i\}$ is the orthonormal basis of R^n. Thus **grad** ($f * \varphi$) is normal to the essential boundary at $\overline{x}_0$, which completes proof of the Lemma 2.2.1. Q.E.D.

Remark 1:

As a consequence of (2.28) we may obtain the formula for the gradient's length $|\mathbf{grad}\ (f * g)|$ at $\overline{x}_0$. I.e., the value of the derivative of $f * g$ in the gradient direction at $\overline{x}_0$ is

$$|\mathbf{grad}\ (f * g)\ (x_0)| = \sqrt{\Big[\int_L [f] \cdot \tilde{\varphi}(\overline{x})\ d\ell\Big]^2 \Big[\sum_{i=1}^{n} \cos^2 (\overline{n}\hat{\ } x_i)\Big]} \qquad (2.29)$$

$$= \int [f] \cdot \tilde{\varphi}(\tilde{x}) d\ell\ .$$

Remark 2:

Conversely, if the magnitude of the regular part of the generalized derivative $\frac{\overline{\partial f}}{\partial x_k}$, namely $\frac{\partial f}{\partial x_k}$, is significant, then each component of the $\nabla(f * g)$ has to be modified by adding a new term equal to $\int_{B_\epsilon} \frac{\partial f}{\partial x_k} \cdot \tilde{\varphi}\ d\overline{x}$ as in (2.25). Thus, the direction of this gradient vector can be *arbitrarily perturbed* from the direction of the normal to the edge. By contrast, if support of the test function B_ϵ is lessened in size in such a way that its n-dimensional volume (i.e., area, if $n = 2$) is made diminutive but its $(n - 1)$ dimensional Hausdorff measure (i.e., length, if $n = 2$) is kept constant, then the conclusions of Lemma 2.2.1 are still valid even if $\frac{\partial f}{\partial x_k}$ is large, as is obviously implied by (2.26). This process of *shrinking the support* of $\varphi(\overline{x})$ without altering its $(n - 1)$ dimensional measure is an exact analogy to the "thin masks" defined in §2.3 during the discussion of the strategy for edge detection in one-dimensional signals.

There is a considerable spectrum of opinions in today's edge detection literature as to how big the detector should be. Some researchers reason that since detection is based on the "differences of averages," masks with wide supports are needed. Others feel that widening the masks will somehow bring into play effects of nonlocal noise, etc. There is also a recent third view that information should be derived from the convolutions with detectors of various widths and then integrated together into the decision procedure. With regard to the last see, [Wit.] for the development of so called *scale space* approach.

While I feel that the ultimate methods of computing singularities are yet to come, possibly by formulating some specialized *free boundary* nonlinear differential or integral equations, perhaps by some clever use of variational principles, and I shall formulate a simple variational setting for an edge detection schema; the analysis and solution estimates of these equations will come from what is called *regularity theory of weak solutions*. We eventually will do away with convolutions, mask detectors, linear filters, etc., but the linear *local* analysis developed in this work will serve as a valuable tool for making theoretical estimates and possibly accommodating *linearization* of this future nonlinear framework.

It is therefore clear that the performance analysis of linear detectors is a worthwhile enterprise.

Here are a few comments on the "proper width" of the edge detector as seen from the FTED (2.2) and Remark 2 of Lemma 2.2.1.

Intuition of the first group, that the edge detection process is founded on the "differences of averages," is exactly right and naturally corresponds to the notion of *generalized differentiation* applied to the discontinuous functions. The principal tool of the distributional approach is to view the world through the "averaging glass" provided by the space of smooth functions $\mathcal{K}_n$. Obviously, all smooth functions φ in $\mathcal{K}_n$ have some finite "width," i.e., n-dimensional Lebesgue measure $\lambda(\text{supp}(\varphi)) \neq 0$. Then the FTED tells that to get a meaningful contribution from jump singularity, the averaging process is not enough—it should be followed by differentiation, thus resulting in the above "differences of averages." Nonetheless, the width of the support of the test function may be arbitrarily small (but not zero) as long as the singularity lies within. This brings the point of the second group, mentioned above. Actually, their premise is fully substantiated by Remark 2 of Lemma 2.2.1. The contribution of the singularity in (2.25) can be hopelessly smeared by the large value of $\int \frac{\partial f}{\partial x_k} \cdot \varphi \cdot d\overline{x}$ resulting from the interaction of the test function with "nonsingular" parts of the function $f(\overline{x})$ within B_ϵ, possibly incorporating the contribution of the noise process in the neighborhood B_ϵ. Observe that even though a noise comprises a singular function, it is not at all the same kind of animal as an essential singularity of Definition 2.4.

The noise singularities, put together, do not constitute a surface distribution; therein lies the essence of the distinction between the "edge" and the noise.

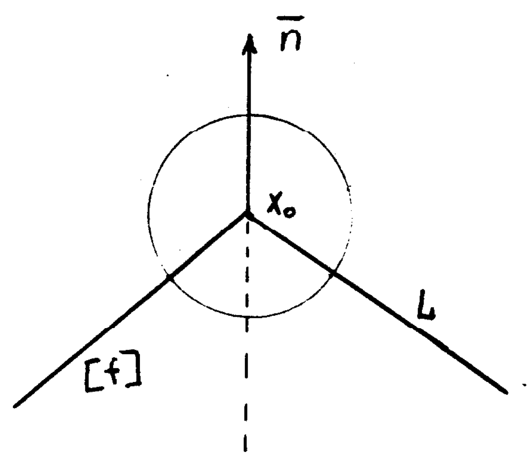
To this end, the use of as narrow as possible masks is strongly indicated. Of course, in the digital imaging setting, the minimal width is determined by the sampling resolution of the image as well as by the signal-to-noise ratios (see [Can.] with regard to the noise estimation for edge detectors).

At this time I shall withhold any comments as to the scale space approach of [Vit.].

We now establish the direction of the generalized gradient for the edge of nonzero curvature, i.e., the nonstraight edge within the support of the detecting mask. We shall see that more assumptions about $f(\bar{x})$ and its behavior on the boundary have to be made in order to get to the conclusions similar to that of Lemma 2.2.1. In the next Lemma 2.2.2 our objective is to evaluate the direction of the gradient vector at the *corner* point of the edge of Figure 2.5:

__Lemma 2.2.2.__

Let $F(\bar{x})$ and $\varphi(\bar{x})$ satisfy the same conditions as in Lemma 2.2.1 and let essential singularity L be a "corner" α of Figure 2.5 and $\overline{x_0}$ be its vertex. Then $\nabla(f(\bar{x}) * \varphi(\bar{x}))\ (\bar{x}_0)$ is in the direction of the bisector of the corner's angle α.

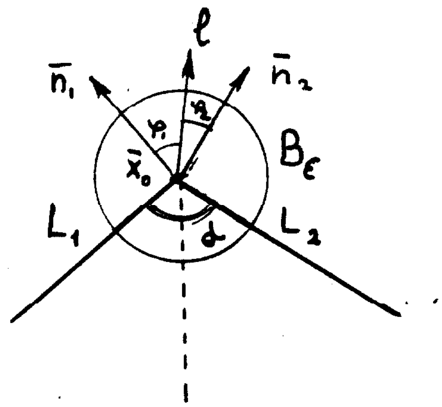


(Fig. 2.5)

__Proof:__

We shall use the fact that the gradient vector shows the direction in which the function is changing the fastest. In order to prove Lemma 2.2.2 we will compute the directional derivative $\frac{\partial}{\partial \ell}\ (f * \varphi)\ (\bar{x}_0)$ in the arbitrary direction ℓ and then determine that it is maximal when it is bisecting α, in half, for this will imply the statement of Lemma 2.2.2.

Let us choose the direction ℓ as depicted in Figure 2.6, where $\overline{n}_1$ and $\overline{n}_2$ are normals to L_1 and L_2 correspondingly, and $\varphi_i = \widehat{n_i\ell}$, $i = 1, 2$. L_1 and L_2 are segments of the boundary within $\text{supp}(\varphi(\overline{x})) = B_\epsilon$.



(Fig. 2.6)

By the FTED and (2.23)

$$A = \left\langle \overline{\frac{\partial f(\overline{x})}{\partial \ell}}, \varphi(\overline{x}) \right\rangle = \int_{B_\epsilon} \frac{\partial f(\overline{x})}{\partial \ell} \cdot \varphi(\overline{x})\, d\overline{x} + \int_L \cos(\widehat{\overline{n}\ell}) \cdot \varphi(\overline{x})\, d\ell \,. \quad (2.30)$$

Recall from the proof of Lemma 2.2.1 that under the above hypothesis on $f(\overline{x})$ and $\varphi(\overline{x})$ we have $\int_{B_\epsilon} \frac{\partial f}{\partial \ell} \cdot \varphi\, d\overline{x} \approx 0$, as was demonstrated by (2.26).

If one ignores the terms which are small, we can continue (2.30) as follows:

$$= \sum_{i=1}^{2} \int_{L_i} \cos(\widehat{\overline{n}_i\ell}) \cdot [f]\, \varphi(\overline{x}) d\ell \,.$$

Now suppose $[f]$ is the same on L_1 and L_2, and $\varphi(\overline{x})$ is circulary and symmetric, then

$$= (\cos(\varphi_1) + \cos(\varphi_2)) \cdot \int_{L_1} [f]\, \varphi(\overline{x}) d\ell \,.$$

Note that $\varphi_2 = \pi - (\alpha + \varphi_1)$, then continue (2.30) as

$$= (\cos(\varphi_1) - \cos(\alpha + \varphi_1)) \cdot \int_{L_1} [f] \cdot \varphi(\bar{x})\, d\ell$$

$$= C \cdot (\cos(\varphi_1) - \cos(\alpha + \varphi_1)) \,.$$

Let φ_1 be a parameter in the above equation. Then differentiating with respect to this parameter gives

$$A' = 2 \sin(\alpha/2) \cdot \cos(\alpha/2 + \varphi_1) = 0 \,. \tag{2.31}$$

Then since $\alpha \neq 0$, $\alpha/2 + \varphi_1 = \pi/2$, from Figure 2.6 we may conclude that ℓ bisects angle α in half if A is to reach extremum among all possible directions of differentiation. A simple differentiation of (2.31) shows that in fact $A'' < 0$; therefore we have *maximum*. This completes the proof of Lemma 2.2.2. Q.E.D.

<u>Remark</u>:

The statement of Lemma 2.2.2 holds true only under additional assumptions of the symmetric behavior of the jump function $[f(\bar{x})]$ with respect to the sides L_1 and L_2 of the "corner" singularity in the vicinity of its vertex $\bar{x}_0$, as well as the circularly symmetric character of the detecting test function $\varphi(\bar{x})$. Observe that if $[f]$ behaves differently on the branches L_1 and L_2, say, the discontinuity jump on L_1 is constant and is smaller than the constant jump across L_2, then similarly to the proof given above, it is shown that the direction of the gradient at $\bar{x}_0$ will have a strong bias towards normal $\overline{n_2}$ of L_2.

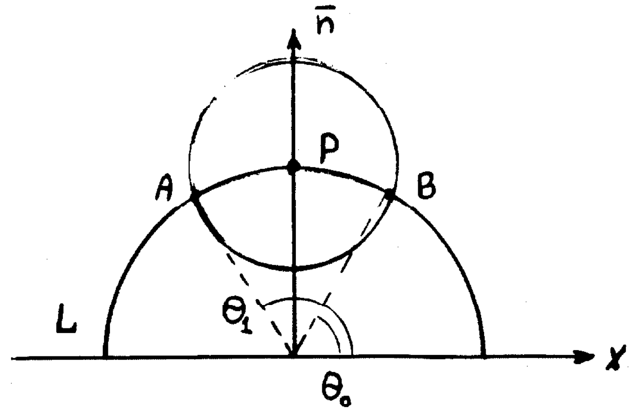
Moreover, we may ask the question of how gradient vector behaves if the center of the mask is shifted along L_i in the neighborhood of the vertex $\overline{x_0}$. Supposing again symmetric conduct of [f], it is easy to verify that as $\overline{x_0}$ is moved along L_i towards the vertex and crosses it, the direction of the gradient at $\overline{x_0}$ *continuously* rotates from the direction of the normal $\overline{n_1}$, to the direction of the normal $\overline{n_2}$. Thus we see that Lemmas 2.2.1 and 2.2.2 are merely referring to the limiting cases when either only one L_i is within supp($\varphi(\overline{x})$), or the vertex of the L_i's coincides with the center of the supp($\varphi(\overline{x})$).

Finally, an instance of the boundary of constant, *nonzero* curvature should be considered.

__Lemma 2.2.3.__

Under the assumptions of Lemma 2.2.2, let L locally be a *circular* essential boundary and $\overline{x_0} \in L$. Then $\nabla(f(\overline{x}) * \varphi(\overline{x}))(\overline{x}_0)$ is normal to the arc L at $\overline{x_0}$.

<u>**Proof:**</u>



(Fig. 2.7)

Let us choose the coordinate system in such a way that the origin lies in the center of curvature O of the portion of L within supp($\varphi(\bar{x})$). (It can be done since L is locally a circular arc), and the y-axis contains $\bar{x}_0$, denoted by P on Figure 2.7.

Let r be the local radius of L and θ its polar angle. Then L is locally given in parametric form

$$x = r \cdot \cos(\theta), \quad y = r \cdot \sin(\theta),$$

and its unit normal $\bar{n}$ is given by the formula

$$\bar{n} = - \frac{(\ddot{x}, \ddot{y})}{|(\ddot{x}, \ddot{y})|} = (\cos(\theta), \sin(\theta)) \ . \tag{2.32}$$

By the FTED (2.23) and discarding the regular contributions accordingly with our assumptions we have

$$\nabla(f(\bar{x}) * \varphi(\bar{x}))(\bar{x}_0) = \left(\int_L n_x \cdot [f] \cdot \varphi(\bar{x}) \cdot d\ell, \int_L n_y \cdot [f] \cdot \varphi(\bar{x}) \cdot d\ell \right), \tag{2.33}$$

where $\overline{n} = (n_x, n_y)$. We replace the line integrals in (2.33) by the definite integrals with respect to the (r, θ) parameters and apply (2.32) to obtain

$$= r \cdot \left(\int_{\theta_0}^{\theta_1} \cos(\theta) \cdot [f] \cdot \varphi(\overline{x}) \cdot d\theta, \int_{\theta_0}^{\theta_1} \sin(\theta) \cdot [f] \cdot \varphi(\overline{x}) \cdot d\theta \right),$$

where θ_0 and θ_1 are as on Fig. 2.7. Because of the symmetry assumptions on $\varphi(\overline{x})$ and $[f]$ and the chosen set-up for the coordinate system and in view of the fact that $\cos(\theta)$ will change its sign across the point P (Fig. 2.7), it follows that $\cos(\theta) \cdot [f] \cdot \varphi(\overline{x})$ is an *odd* function along the arc APB. Thus we may write

$$\int_{\theta_0}^{\theta_1} \cos(\theta) \cdot [f] \cdot \varphi(\overline{x}) \cdot d\theta = 0. \tag{2.34}$$

Substituting this estimate into (2.33) we get

$$\nabla(f(\overline{x}) * \varphi(\overline{x}))\,(\overline{x}_0) = (0, C), \tag{2.35}$$

where constant $C = r \int_{\theta_0}^{\theta_1} \sin(\theta) \cdot [f] \cdot \varphi(\overline{x}) \cdot d\theta$, which is not zero since $\sin(\theta)$ is even across point P.

But (2.35) implies that the examined gradient vector is in the direction of the y-axis which is normal to the arc L at $\overline{x_0}$ by its very construction. Hence $\nabla(f(\overline{x}) * \varphi(\overline{x}))\,(\overline{x}_0)$ is normal to L. Q.E.D.

<u>**Remark:**</u>

Lemma 2.2.3 may be extended to more general curves as long as local symmetries within supp(φ) of the foregoing discussion are preserved.

Finally, a few comments on the usefulness of the above analysis and a brief summary of what we have learned.

Edge detection schemes of today are built to measure certain quantities in the direction provided by the estimated gradient of the image. Based on these measurements, higher-order procedures are devised to accept the point as belonging to the edge or to reject it. Clearly, such a decision rests on the premise that the gradient *is* normal to the potential edge. No analysis to substantiate such a claim has ever been given. The purpose of this paragraph is to fill the above-mentioned gap and to demonstrate the analytical power of the quite innocent-looking FTED. Meanwhile, an insight into the nature of the dependency of the convolution product's behavior in the vicinity of jump singularity on the geometry of this singularity, was acquired. We have learned that intuition can be very misleading in neighborhoods with large curvatures (of the edge) or when the boundary itself is singular, e.g., corners. We are now in a position to state and prove the main generic facts about various *local* feature detectors commonly employed in the industrial applications of today.

§2.5. *First look at how to detect edges and why existing schemes succeed or fail.*

This paragraph is not intended to be a review of the existing edge detecting techniques. An excellent survey may be found in [Bli.]. Rather, we would like to take a look at what actually has been gained by the introduction of the Fundamental Theorem of Edge Detection. Of course we have seen how useful FTED is in analyzing the local structure of convolution products when one of the integrands has jump singularities. Now we shall relate the newly acquired machinery to the actual process of detecting singularities in images. It should be clear that whenever the word 'image' is invoked, it is not confined to the domain of $BV(\Omega)$ functions of two variables. Obtaining mathematical generality is not our goal either. However, applications demand a much broader concept of the 'image.' Specifically, analysis of *moving* imagery demands consideration of its three-dimensional nature. Dimension is further increased when stereo motion sequences are considered. Color spectrum contributes to additional dimensionality.

The most important consequence of the FTED is that the contribution of the *singularity* to the inner product with a smooth function is

$$<f, g>_{sing} = \int_L \cos(\overline{n}\ x_i) \cdot [f] \cdot G(\overline{x}) \cdot d\ell \ , \tag{2.36}$$

where $\frac{\partial G(\overline{x})}{\partial x_i} = -g(\overline{x})$.

Let us assume that $G(\overline{x})$ is circularly symmetric, whose center $\overline{x_0}$ is right on top of the singularity L. Then one way to increase the singular contribution in (2.36) is to rotate the direction of differentiation so that $<f, g>_{sing}$ reaches its maximum. In order to see the effect of rotation, assume the portion of the boundary within $supp(G(\overline{x}))$ to be a straight segment. Then

$$<f, g>_{sing} = \cos(\overline{n}\hat{\ } x_i) \cdot \int_L [f] \cdot G(\overline{x}) \cdot d\ell ,$$

and according to (2.29)

$$= \cos(\overline{n}\hat{\ } x_i) \cdot \left|\mathbf{grad}\ (f * G)\ (\overline{x_0})\right| . \tag{2.37}$$

Clearly the magnitude of the gradient is independent of the rotation angle and $\cos(\overline{n}\hat{\ } x_i)$ is maximal when the direction of differentiation coincides with the normal $\overline{n}$ at $\overline{x_0}$. Then the grand strategy is to maximize the contribution (2.37) of the jump singularity by rotating the direction of differentiation until it points exactly parallel to the normal of the edge. If this contribution is "significant," $\overline{x_0}$ is declared to be the *edge point*. The trouble starts with deciding on the magnitude of the threshold—separating significant contributions from the rest of the measurements. To illustrate the difficulty we next consider

$$f_\epsilon\ (x, y) = M_\epsilon \cdot \sqrt{x^2 + y^2} ,$$

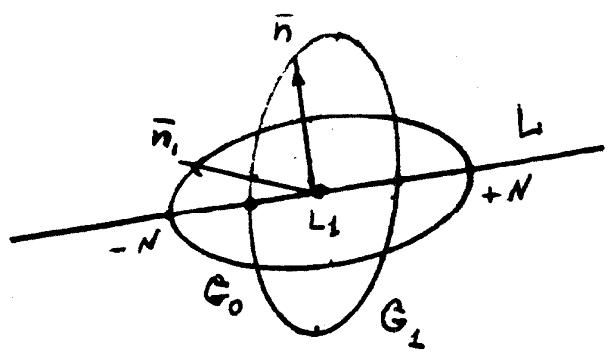
where M_ϵ is some constant.

Next compute

$$|\nabla\ (f_\epsilon * G)|\ (x_0, y_0) = M_\epsilon \cdot \iint_{B_\epsilon} G \cdot dxdy, \tag{2.38}$$

where B_ϵ is a circular support of G. This suggests that by making M_ϵ arbitrarily large, $|\nabla(f_\epsilon * G)|$ can be made to exceed any *fixed* threshold, since $\iint_{B_\epsilon} G\ dxdy$ is bounded. Yet $f_\epsilon(x, y)$ is a smooth function lacking any singularities. The "narrow masks" strategy is employed here to overcome the interference with "smooth noise."

The circularly supported detector G is replaced by a test function G with an elongated support $\mathrm{supp}(G) = B(\epsilon, N) = \{x, y \mid |x| \leq \epsilon, |y| \leq N\}$, and $\epsilon \ll N$, so that $\lim\limits_{\epsilon \to 0} \iint\limits_{B(\epsilon, N)} G \, dxdy = 0$. By choosing sufficiently small ϵ, the regular contribution (2.38) is rendered harmless. However, singular contribution, if it is present, is still measurable by differentiating the elongated mask in the direction of its large axis. Figure 2.8 illustrates this process for two rotated positions of the mask.



(Fig. 2.8)

If the long axis of G is aligned with the singular boundary L as on Fig. 2.8,

$$\left\langle f, \frac{\partial}{\partial \bar{n}} G_0 \right\rangle = |\nabla(f * G_0)| \, (\bar{x}_0) \, ,$$

and from the shape of $B(\epsilon, N)$ we get

$$= \int\limits_{[-N,\, N]} [f] \cdot G_0 \cdot d\ell + O(\epsilon) \, , \tag{2.39}$$

where [-N, N] denotes the segment of the singular boundary within supp(G_0). While for G_1 we have

$$\left\langle f, \frac{\partial}{\partial \bar{n}_1} G_1 \right\rangle = \int_{L_1} \bar{n}_1 \bullet \bar{n} \cdot [f] \cdot G_1 \cdot d\ell + O(\epsilon) \,, \qquad (2.40)$$

where $\bar{n}_1$ is a unit normal to the larger axis of G_1 and L_1 is a segment of L within supp (G_1).

We next observe that if ϵ is small, the line integral in (2.40) yields insignificant value since both the length of L_1 and $\cos(\bar{n}_1 \hat{} \bar{n})$ are small. But the line integral in (2.39) depends only on the length 2N of the larger axis, i.e., is independent of ϵ.

From (2.39) and (2.40) it also follows that we could measure either the value of the directed derivative or compute the gradient's magnitude. If the mask is perfectly aligned (as G_0 on Fig. 2.8), both measurements yield the same value. On the other hand, different values are obtained for the rest of the rotations. They differ by the presence of the cos $(\bar{n}_1 \hat{} \bar{n})$ term in (2.40), since $|\nabla(f * G_1)|\ (\bar{x}_0)$ depends only on the length of L_1. Therefore, the *directional derivative measurement reveals some extra information* about the relative location of the singularity.

The above strategy seems to rescue the thresholding technique from the "smooth noise." There is a second reason, however, why it is not a practical solution.

Again, assuming the simplest case of straight edge, we see that the singular contribution is proportional to $\int_L [f] \cdot d\ell$. According to (1.1) of the edge formation process, the size of the jump [f] is determined by the reflectivity function and by the luminance of the scene. Hence, under the single threshold, *measurements from darker regions may be rejected even if the jump singularity is residing there.* We shall return to this question later in the description of the "simple edge detector."

Now we take a brief look at several popular edge detectors, introduced in §2.3.

Roberts' masks R_1 and R_2 compute two-point finite differences in two mutually perpendicular directions given by the diagonals of the mask. Therefore $\sqrt{R_1^2 + R_2^2}$ is a digital approximation of the magnitude of the gradient. Assuming that the imaging process has introduced some smoothing of the original picture, we may expect $\sqrt{R_1^2 + R_2^2}$ to be an approximation to $\int_{L_\epsilon} [f]\, \varphi(\overline{x}) \cdot d\overline{x}$, where L_ϵ is a segment of the edge within the support of the "smoothing filter" of the imaging process. This detector is known to be useful for noise-free images, including "smooth noise." Thus is not surprising, since the above integral estimate is valid. However, once the smoothing assumption is not sound, e.g., when white noise is introduced, the previous integral estimate is not valid. Not surprisingly, experiments with the Roberts detector show it to be worthless under noisy conditions.

Analysis of the Sobel's and "compass gradient" detectors of §2.3 proceeds in analogous fashion. We observe that Sobel's vertical mask is a 3×3 restriction of the central difference operator Δ_x applied to the

$$\tilde{S}_{\text{vertical}} = \begin{bmatrix} 0 & x_1 & 2 & x_1 & 0 \\ 0 & x_2 & 4 & x_2 & 0 \\ 0 & x_3 & 2 & x_3 & 0 \end{bmatrix},$$

where x_i are any numbers, and $\tilde{S}_{\text{horizontal}}$ is its rotated version. Thus Sobel's detection is equivalent to convolving with axis symmetric masks $\tilde{S}_{vertical}$ and $\tilde{S}_{horizontal}$ and then differentiating in the horizontal and vertical directions correspondingly. Hence, it is just an approximation of generalized differentiation. Similarly the "compass gradient" detector is a "poor man's" version of the rotating test function detector described previously.

In both cases the support of approximation is a 3×3 square, making it susceptible to white noise.

The reduction of the noise sensitivity is usually achieved by widening the mask's support (see differences of averages in [Ros. Kak.]). Again, nothing new and exciting here, just an ad hoc approximation of the generalized differentiation.

In sum, the simple detecting algorithm proposed here has all advantages of the well known schemas, while it does not inherit many of their shortcomings. For it is formulated on the basis of solid analysis provided by the FTED.

At this point the reader may disappointedly ask if this is the ultimate singularity detection theory? And how about the promise of a *totally new* beginning with such a glorious name—Numerical Analysis of Singularities. Of course, the FTED is a very useful tool, as was demonstrated in numerous calculations and lemmas. It helps to explicate many existing algorithms with mathematical rigor and guides the design of a much cleaner version. Yes, the proposed scheme is mathematically sound. But frankly, it is not *operationally* much different from what is presently the most successful edge detection algorithm of [Can.], derived without any use of generalized calculus. In fact, if this were all we were potentially able to build, the author would not bother with taking so many mathematical precautions. And definitely, he would not prophesize the unfolding of any new field!

And what about all the other kinds of detectors? We cannot ignore such popular schemes as zero-crossings, line detectors, spot detectors, corner detectors, motion detectors, etc. I can almost hear the reader's voice,—"Oh no, not FTED again!"

§2.6. *Marr's zero-crossings of Laplacian and Laplacian line detection.*

Much has been written on the "Theory of edge detection" of D. Marr [Mar.]. The discussion of its merits as a theory of human vision is beyond the modest task of the author to evaluate Marr's zero-crossings performance as a singularity detector. This scheme is a relatively recent approach, compared to the schemes already considered.

The original idea in [Ma. Hi.1] was "that sudden intensity change will give rise to a peak or trough in the *first derivative* or, equivalently, to a *zero-crossing* in the *second derivative*..." While it is never defined what exactly is meant by the "sudden intensity change" and what kind of "first derivative" the authors had in mind, the "second derivative" is chosen to be the Laplacian $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

Actually, the scheme works by finding *zeros* of $\nabla^2(G(x, y) * I(x, y))$, where $G(x, y)$ is a two-dimensional Gaussian and $I(x, y)$ is an image function. In what follows this expression will be referred to as Marr's detector.

We immediately observe the explicit presence of the integration with the smooth function in Marr's detector. Of course, $G(x, y) \notin \mathcal{K}_2$, since it does not vanish outside any finite support. However $G(x, y) =$

$C \cdot \exp(-(x^2 + y^2)/2\delta^2)$ decreases at infinity, together with all its derivatives, faster than any power of $(x^2 + y^2)^{-1/2}$. Then G(x, y) belongs to the space of so called 'slow growth' functions. This spaces occupies the role almost identical to $\mathcal{K}_n$ in previous considerations. This means that *generalized functions* of slow growth, or 'tempered distributions,' are definable in the way analogous to $\mathcal{K}_n^1$. Throughout this work, Δ will denote the Laplacian operator. Consequently, Marr's detector is written as

$$\Delta\ (f(\overline{x}) * G(\overline{x})) \tag{2.41}$$

It follows that Marr's detector (2.41) is just a linear functional for each shift of the $G(\overline{x})$, i.e., a tempered distribution. Since in practice only functions with bounded support are used, without loss of generality we may assume $G(\overline{x})$ to be one of our familiar cup-shaped test functions. But then (2.41) is a second-order differential operator applied to the distribution in $\mathcal{K}_n'$.

Since we want to evaluate behavior of the Marr's detector in the vicinity of singularities, it is crucial to be able to decompose (2.41) in such a way that the *singular* contribution of $f(\overline{x})$ may be estimated. Clearly, we want to carry on with Laplacian what was already performed with the first order partial derivatives in (2.23) of FTED. The brute force method would proceed by using the rule of differentiating the convolution

$$\Delta(f(\overline{x})) * G(\overline{x})) = (\overline{\Delta}(f(\overline{x}) * G(\overline{x}))\ , \tag{2.42}$$

where $\overline{\Delta}$ is *generalized* Laplacian, i.e.

$$\bar{\Delta} = \overline{\frac{\partial^2}{\partial x^2}} + \overline{\frac{\partial^2}{\partial y^2}} = \frac{\bar{\partial}}{\partial x}\left[\frac{\bar{\partial}}{\partial x}\right] + \frac{\bar{\partial}}{\partial y}\left[\frac{\bar{\partial}}{\partial y}\right] \tag{2.43}$$

in the sense of Definition 2.5.

Next we may attempt to compute $\bar{\Delta}(f(\bar{x}))$ by applying FTED (2.23) in (2.43) and compute individual $\frac{\bar{\partial}}{\partial x_i}\left[\frac{\bar{\partial}}{\partial x_i}\right]$. However a look at (2.23) shows the potential problem—we have no rules yet to compute

$$\frac{\bar{\partial}}{\partial x_i}\left([f]_{L_k} \cdot \delta_{L_k}\right) . \tag{2.44}$$

What is especially notable about (2.44) is that rules are needed for generalized differentiation of the surface distribution, multiplied by the density function defined only on this surface. The elements of the distribution theory introduced so far have nothing in their arsenal to compute (2.44). Actually, the classical distribution theory does not offer any remedy in (2.44). Eventually we will have to come back to (2.44) and solve it. However in the case of $\bar{\Delta}$ we can bypass (2.44) and compute the generalized Laplacian without resorting to the brute force method.

<u>**Lemma 2.6.1**</u>.

Let both $f(\bar{x}) \in BV(\Omega)$ and the regular part $\frac{\partial f}{\partial x_i} \in BV(\Omega)$ be piecewise smooth functions with essential boundaries L_K and M_K correspondingly, and with the rest of the notation being as in FTED (2.2). Then one has

$$\bar{\Delta} f = \Delta f + \sum_i \left[\frac{\partial f}{\partial n}\right]_{M_i} \cdot \delta_{M_i} + \sum_K \frac{\bar{\partial}}{\partial n}\left([f]_{L_k} \cdot \delta_{L_K}\right) , \tag{2.45}$$

where $\left[\frac{\partial f}{\partial n}\right]_{M_i}$ is a size of the jump of the f's derivative $\left.\frac{\partial f}{\partial n}\right|_{M_i}$ taken in the direction of the outer normal to M_i on M_i, and $[f]_{L_k}$ as in the previous convention.

Proof:

We shall apply a well known Green's identity for the Laplacian operator (see [Mars. Tr.]), which states that

$$\int\limits_{(\mathfrak{D})}\!\!\dots\!\int (v\Delta u - u\Delta v) \cdot d\bar{x} = \int\limits_{L} \left(v \frac{\partial u}{\partial n} - u \frac{\partial v}{\partial n}\right) \cdot d\ell\,, \tag{2.46}$$

where $\bar{x} = (x_1, x_2, \dots, x_n)$, closed set $\mathfrak{D} \subset R^n$, and L is a boundary of $\mathfrak{D}$.

Let us now suppose that $\varphi(\bar{x}) \in \mathcal{K}_n$ and $f(\bar{x}) \equiv 0$ for $x \in R^n/\mathfrak{D}$. We may further suppose that f_L^+, f_L^-, $\left(\frac{\partial f}{\partial n}\right)_L^{\pm}$ are correspondingly limits of outside and inside boundary values of $f(\bar{x})$ and $\frac{\partial f}{\partial n}$ on L (i.e., traces of f and $\left(\frac{\partial f}{\partial n}\right)$ on L).

We now deduce with help of Green's identity (2.46)

$$\begin{aligned}\int\limits_{(\mathfrak{D})}\!\!\dots\!\int f(\bar{x}) \cdot \Delta\varphi(\bar{x}) \cdot d\bar{x} &= \int\limits_{(\mathfrak{D})}\!\!\dots\!\int \Delta f(\bar{x}) \cdot \varphi(\bar{x}) \cdot d\bar{x} \,+ \\ &+ \int\limits_{L} f_L^- \cdot \frac{\partial\varphi(\bar{x})}{\partial n} \cdot d\ell - \int\limits_{L} \left(\frac{\partial f}{\partial n}\right)_L^- \cdot \varphi(\bar{x}) \cdot d\ell\,.\end{aligned} \tag{2.47}$$

And similarly if we set $f(\bar{x}) \equiv 0$ for $x \in \mathfrak{D}$, then we deduce

$$\begin{aligned}\int\limits_{R^n\backslash\mathfrak{D}}\!\!\dots\!\int f(\bar{x}) \cdot \Delta\varphi(\bar{x}) \cdot d\bar{x} &= \int\limits_{R^n\backslash\mathfrak{D}}\!\!\dots\!\int \Delta f(\bar{x}) \cdot \varphi(\bar{x}) \cdot d\bar{x} \,- \\ &- \int\limits_{L} f_L^+ \frac{\partial\varphi(\bar{x})}{\partial n} \cdot d\ell + \int\limits_{L} \left(\frac{\partial f}{\partial n}\right)_L^+ \cdot \varphi(\bar{x}) \cdot d\,\ell\,.\end{aligned} \tag{2.48}$$

Notice different signs for line integrals in (2.47) and (2.48) which result from taking the *same direction of the normal* in two opposite applications of (2.46). Adding (2.47) and (2.48) gives us for $f(\overline{x})$ with nonzero values inside and outside $\mathfrak{D}$,

$$\int_{R^n}\!\!\dots\!\int f(\overline{x}) \cdot \Delta\, \varphi(\overline{x}) \cdot d\overline{x} = \int_{R^n}\!\!\dots\!\int \Delta f(\overline{x}) \cdot \varphi(\overline{x}) \cdot d\overline{x} \; -$$

$$- \int_L (f_L^+ - f_L^-) \cdot \frac{\partial\varphi(\overline{x})}{\partial n} + \int_L \left(\left(\frac{\partial f}{\partial n}\right)_L^+ - \left(\frac{\partial f}{\partial n}\right)_L^-\right) \cdot \varphi(\overline{x}) \cdot d\ell \; .$$

Using our notation for jumps gives

$$\begin{aligned} &= \int_{R^n}\!\!\dots\!\int \Delta f(\overline{x}) \cdot \varphi(\overline{x}) \cdot d\overline{x} \; - \\ &- \int_L [f]_L \cdot \frac{\partial\varphi(\overline{x})}{\partial n} \cdot d\ell + \int_L \left[\frac{\partial f}{\partial n}\right]_L \cdot \varphi(\overline{x}) \cdot d\ell \; . \end{aligned} \tag{2.49}$$

In view of the identity

$$<\overline{\Delta} f, \varphi> \; = \; <f, \Delta\varphi> \tag{2.50}$$

and definition formula (2.14) of *double layer*, (2.49) is rewritten as follows:

$$<\overline{\Delta} f, \varphi> \; = \; <\Delta, f, \varphi> + \left\langle \left[\frac{\partial f}{\partial n}\right]_L \cdot \delta_L, \varphi \right\rangle + \left\langle \frac{\partial}{\partial n} ([f]_L \cdot \delta_L), \varphi \right\rangle . \tag{2.51}$$

Observe that the sign in front of $\frac{\partial}{\partial n}([f]_L \cdot \delta_L)$ is $+$ because of the $(-1)^K$ multiplier in the (2.15) definition of k-th multilayer ($k = 1$ for double layer).

In light of the definition of equality of generalized functions and observing that $\varphi(\bar{x})$ is an arbitrary test function in (2.51) we have

$$\bar{\Delta} f = \Delta f + \left[\frac{\partial f}{\partial n}\right]_L \cdot \delta_L + \frac{\partial}{\partial n}([f]_L \cdot \delta_L) . \tag{2.52}$$

Now, we combine boundaries, across which the jump of the normal derivative $\left[\frac{\partial f}{\partial n}\right]_{M_i}$ is not identically zero, into the set of essential boundaries M_i. Similarly, for the set L_K of essential boundaries across which $[f]_{L_k} \not\equiv 0$. Observe that some boundaries L will fall in both sets, if both f(x) and its normal derivative have a nonzero jump across L. Also note that it is possible to have a boundary across which *only one* of the quantities above undergoes a jump, e.g., the function $f(\bar{x})$ may be essentially discontinuous on the boundary L but its normal derivative across L is not in the sense that $\left[\frac{\partial f}{\partial n}\right]_L \equiv 0$. This substantiates remarks made in §2.1 on the possibility of quite complicated local structure of singularities.

Finally, using the induction argument on the number of boundaries M_i and L_i and (2.52), we arrive at the statement (2.45) of Lemma 2.6.1. Q.E.D.

<u>Remark:</u>

When repeating induction in (2.52) one can get the mistaken impression that summing up regular Δf should result in different regular part of (2.45). It is not so if one follows the step of the proof above where domain R^n is subdivided each time in two *disjoint* areas. Clearly, the proof of Lemma 2.6.1 is still valid if boundaries intersect each other. The essential meaning of the above lemma and (2.45) is stressed below. The Generalized Laplacian of $f(\bar{x})$ is *uniquely* decomposable into the sum of three following components:

(i) The regular part, which is just a regular function.

(ii) The sum of *single layers* with density equal to the *variable size of the jump of normal derivative* of $f(\overline{x})$ across certain rectifiable curves.

(iii) The sum of *double layers* with density equal to the variable *size of the jump of* $f(\overline{x})$ across some rectifiable curves.

Looking at formula (2.45), we may easily analyze the performance of the Marr's zero-crossing detector.

<u>Corollary</u> <u>2.6.1</u> (Genesis of Marr's detector)

Let $f(\overline{x})$ be a smooth function with essential singular boundary L and let $\varphi(\overline{x}) \in \mathcal{K}_n$ be a circularly symmetric test function. Then (2.45) yields

(i) If L is of zero curvature and $\overline{x_0} \in L$, then

$$\Delta(f(\overline{x}) * \varphi(\overline{x}))\,(\overline{x_0}) = 0\ .$$

if either $\mathrm{supp}(\varphi(\overline{x}))$ or regular part $\Delta f(\overline{x})$ is negligently small, and $\left[\frac{\partial f}{\partial n}\right]_L \equiv 0$.

(ii) If L is a "corner" and $\overline{x_0}$ is its vertex, then under similar assumptions as in (i) one has

$$\Delta(f(\overline{x}) * \varphi(\overline{x}))\,(\overline{x_0}) = 0$$

(iii) Under the assumptions of (i), and $\varphi_\alpha(\overline{x}) \in \mathcal{K}_n$ assumed to be just axis symmetric,

$$\Delta(u * \varphi_\alpha)(\overline{x}_0) = 0 \ ,$$

where $\varphi_\alpha(\overline{x})$ is $\varphi(\overline{x})$ rotated by angle α as on Fig. 2.9. I.e., the *zero-crossing condition alone does not define normal direction of the singularity.*

Proof:

(i) By (2.45) it follows that (see Fig. 2.4)

$$\Delta(f * \varphi)(\overline{x}_0) = \int\limits_{supp(\varphi)-B_\epsilon} \Delta f \cdot \varphi \cdot d\overline{x} + \int\limits_L \left[\frac{\partial f}{\partial n}\right]_L \cdot \varphi \cdot d\ell - \int\limits_L [f]_L \cdot \frac{\partial \varphi}{\partial n} \cdot d\ell$$

For small enough B_ϵ or vanishing *regular* Δf we have

$$= \int\limits_L \left[\frac{\partial f}{\partial n}\right] \cdot \varphi \cdot d\ell - \int\limits_L [f] \cdot \frac{\partial \varphi}{\partial n} \cdot d\ell \ .$$

The identity $\left[\frac{\partial f}{\partial n}\right] \equiv 0$ implies that

$$= -\int\limits_L [f] \cdot \frac{\partial \varphi}{\partial n} \cdot d\ell \ . \qquad (2.53)$$

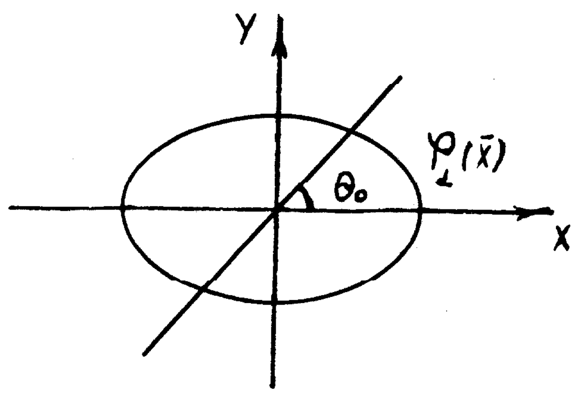
However, $\varphi(\overline{x})$ is a circularly symmetric test function, whose center $\overline{x_0}$ is on the hyperplane L, since L is of zero curvature. Thus we may write $\left.\frac{\partial \varphi}{\partial n}\right|_L \equiv 0$ and so the result follows from (2.53)

$$\Delta(f * \varphi)(\overline{x_0}) = 0 \quad \text{Q.E.D.}$$

Remark:

To satisfy the zero-crossing condition of Marr's detection procedure, a horde of extra conditions had to be imposed on the function $f(\bar{x})$. This is in spite of the simplicity of the singularity L. Armed with an *exact* estimate (2.45) of the singular contribution into the Laplacian of convolution product, we clearly see how *poorly* the zero-crossing condition of Marr's characterizes generic jump singularity. In particular, it is natural to expect that $\left[\frac{\partial f}{\partial n}\right]_L \not\equiv 0$ for singularities corresponding to *occluding contours* in images, as seen on Fig. 1.2. And only *cast shadows* have vanishing size of the jump of the normal derivative across an edge as on Fig. 1.3. Thus the zero-crossing detector will entirely miss any nontrivial occluding boundaries, which are, of course, the most useful kind of singularities, corresponding to the singularities of the visual mapping in §1.1.3.

(ii) To demonstrate zero-crossing property for the "corner" of Fig. 2.5 we can carry the computation (2.53) on each hyperplane of the corner and deduce the required identity by observing that $\left.\frac{\partial\varphi}{\partial n}\right|_{L_i} \equiv 0$ on each side L_i, since $\bar{x}_0$ is the center of symmetry of $\varphi(\bar{x})$. Q.E.D.



(Fig. 2.9)

(iii) Let $\varphi_\alpha(\bar{x})$ be the elongated test function, rotated by some angle α with respect to the singular boundary L as on Fig. 2.9.

The proceeding in (2.53) leads to

$$\Delta\ (f * \varphi)\ (\overline{x_0}) = -\int_L [f] \cdot \frac{\partial \varphi}{\partial n} \cdot d\ell\ . \tag{2.53.1}$$

We shall show that

$$\frac{\partial}{\partial \alpha}\ \varphi(-x, -y) = -\frac{\partial}{\partial \alpha}\ \varphi(x, y)\ , \tag{2.54}$$

where α is an arbitrary direction of differentiation on Fig. 2.9.

Let $x = r \cdot \cos(\theta)$, $y = r \cdot \sin\theta$, then

$$\frac{\partial \varphi}{\partial \alpha} = \cos(\alpha\hat{\ }x) \cdot \frac{\partial \varphi}{\partial x} + \cos(\alpha\hat{\ }y) \cdot \frac{\partial \varphi}{\partial y}\ . \tag{2.55}$$

We also have

$$\frac{\partial \varphi}{\partial x} = \cos(\theta) \cdot \frac{\partial \varphi}{\partial r} - \frac{\sin(\theta)}{r} \cdot \frac{\partial \varphi}{\partial \theta} \tag{2.56}$$

$$\frac{\partial \varphi}{\partial y} = \sin(\theta) \cdot \frac{\partial \varphi}{\partial r} + \frac{\cos(\theta)}{r} \cdot \frac{\partial \varphi}{\partial \theta}\ . \tag{2.57}$$

Thus we obtain from (2.56) and (2.57)

$$\frac{\partial \varphi}{\partial x}\ (r, \theta + \pi) = -\ \frac{\partial \varphi}{\partial x}\ (r, \theta), \tag{2.58}$$

since $\frac{\partial \varphi}{\partial r}\ (r, \theta + \pi) = \frac{\partial \varphi}{\partial r}\ (r, \theta)$ by the symmetry of $\varphi(\overline{x})$ with respect to the origin. Similarly, we get

$$\frac{\partial \varphi}{\partial y}(r, \theta + \pi) = -\frac{\partial \varphi}{\partial y}\ (r, \theta)\ . \tag{2.59}$$

Substituting (2.58) and (2.59) into (2.55) we find

$$\frac{\partial\varphi}{\partial\alpha}(r, \theta + \pi) = -\frac{\partial\varphi}{\partial\alpha}(r, \theta),$$

which can be rewritten as (2.54), i.e.,

$$\frac{\partial\varphi}{\partial\alpha}(-x, -y) = -\frac{\partial\varphi}{\partial\alpha}(x, y) .$$

Returning to (2.53.1) and replacing the line integral with its definite integral representation in terms of parameters (r, θ) we have

$$\Delta(f * \varphi_\alpha)(\overline{x_0}) = \int_0^\infty \frac{\partial\varphi}{\partial n}(r, \theta_o) \cdot [f] \cdot dr +$$

$$+ \int_0^\infty \frac{\partial\varphi}{\partial n}(r, \theta_0 + \pi) \cdot [f] \cdot dr ,$$

which yields $\Delta(f * \varphi_\alpha)(\overline{x_0}) = 0$ upon applying (2.54) and the assumption that $[f]$ is an even function locally on L with respect to the origin $\overline{x_0}$. Q.E.D.

Remark:

Replacing the circularly symmetric mask with the elongated mask in the Laplacian detector preserves the zero-crossing as long as the stringent conditions of (iii) are satisfied. Such a replacement is desirable for the same reason the "thin masks" were introduced in the previous analysis. In fact, making $\text{supp}(\varphi(\overline{x}))$ as on Fig. 2.9 greatly relaxes demands on how small the regular Δf should be, for it can make the area of $\text{supp}(\varphi)$ arbitrarily small without affecting the contributions of singularity. However, the shrinking support of φ does not rescue Marr's detector from its fundamental flow brought up in the remark to (i) above. And the important observation is that contrary to our experience with previous edge detectors, the elongated zero-crossing detector alone yields no directional information about the normal of the edge.

Another useful application of (2.45) is in revealing why the Laplacian operator has peculiar sensitivity to the *pair* of jump singularities positioned alongside of each other, usually designated by the term "*lines*."

So far we were occupied with the case of isolated essential boundary within the support of the test function. However there is no reason why we cannot extend our analysis and *compute interaction of singularities*.

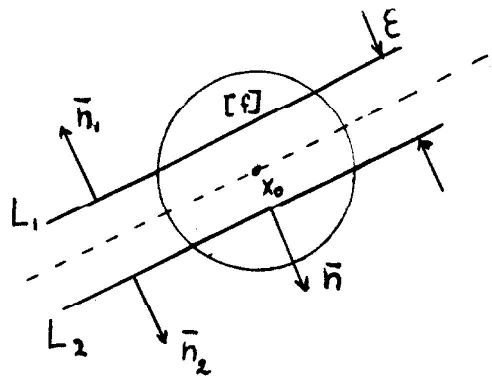
So called "lines" in images provide the simplest example of such interaction.

<u>Corollary</u> <u>2.6.2</u> (Genesis of Laplacian line detector)

Let $f(\overline{x})$, $\varphi(\overline{x})$ be as in Corollary 2.6.1 and L_1, L_2 be two locally ϵ-congruent essential jump singularities, so that for corresponding points $[f]^{\overline{n}}_{L_2} = -[f]^{\overline{n}}_{L_2}$, where jumps are taken in the direction provided by the normal $\overline{n}$ and $\overline{x_0}$ is chosen as on Figure 2.10. Then

(i) $\dfrac{\partial}{\partial \nu} (f * \varphi)(\overline{x_0}) = 0$, where ν is arbitrary direction, (2.60)

(ii) $$\Delta(f * \varphi)(\overline{x_0}) = -2 \int_{L_1} [f]_{L_1} \cdot \frac{\partial \varphi}{\partial n_1} \cdot d\ell \; . \tag{2.61}$$



(Fig. 2.10)

<u>Proof</u>:

(i) By the FTED and assumption that the regular part of $\frac{\overline{\partial f}}{\partial \ell}$ is small (or supp(φ) is small) we have

$$\frac{\partial}{\partial \nu} (f * \varphi)(\overline{x_0}) = \int_{L_1} [f]^{\overline{\Gamma}}_{L_1} \cdot \cos(n^\wedge \nu) \cdot \varphi(\overline{x}) \cdot d\ell + {}$$

$$+ \int_{L_2} [f]^{\overline{\Gamma}}_{L_2} \cdot \cos(n^\wedge \nu) \cdot \varphi(\overline{x}) \cdot d\ell \, .$$

From the symmetries on Fig. 2.10 and a symmetry of the jump conditions we get as promised

$$= 0 \quad \text{Q.E.D.}$$

<u>Remark</u>:

Relation (i) testifies to the fact that locally parallel pairs of singularities cannot be detected by the edge detectors based on first-order generalized differentiation. This includes all the schemas considered so far with the exception of the Laplacian detectors.

(ii) The double layer term in (2.45) deceptively appears to be dependent on the choice of the normal vector $\overline{n}$. In fact, it is not—for $[f]_{L_K}$ jump is computed in the direction of chosen $\overline{n}$. So changing the direction of $\overline{n}$ will also cause change of the sign $[f]_{L_K}$. In particular, we may take $\overline{n_1}$ and $\overline{n_2}$ as on Fig. 2.10, which makes

$$[f]_{L_1}^{\overline{n}_1} = [f]_{L_1} = [f]_{L_2} = [f]_{L_2}^{\overline{n}_2} .$$

Proceeding in the same manner as in (2.53) we get

$$\Delta(f * \varphi)(\overline{x_0}) = -\sum_{i=1}^{2}\left(\int_{L_i} [f]_{L_i} \cdot \frac{\partial\varphi}{\partial \mathbf{n}_i} \cdot d\ell\right) = -2\int_{L_i} [f]_{L_i} \cdot \frac{\partial\varphi}{\partial \mathbf{n}_i} \cdot d\ell .$$

<u>**Remark:**</u>

(2.61) demonstrates that the Laplacian of the convolution product with a smooth test function *aggregates* singular contributions of nearby jump singularities of "line" type, while usual edge detectors *annihilate* this contribution. However, the usual Laplacian mask

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

is a finite-difference approximation of the $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$ operator. If it is applied to a nonsmoothed image or noisy image the above singular contribution will be severely perturbed. Corollary 2.6.2 strongly suggests that the use of directed thin masks in conjunction with the Laplacian operator is much more appropriate. Also notice that the width of the mask $\varphi_\alpha(\bar{x})$ could be chosen so that $\left.\frac{\partial\varphi}{\partial \mathbf{n}}\right|_{L_i}$ is maximal when the mask and the line are perfectly aligned, so the width of the "line" determines the optimal width of the detecting mask.

Yet another application of (2.45) is computing *second-order* singularities which correspond to the loci of single layers $\left[\frac{\partial\varphi}{\partial \mathbf{n}}\right]_{M_i} \cdot \delta_{M_i}$ in (2.45). Observe that this is not a new problem. Indeed, we already proposed a coarse technique for detection of single layers, called rotating thin masks, in §2.5. Of course, it was not explicitly referred to as a "single layer detector," but a quick look at FTED (2.23), which was the principal foundation for the scheme, shows that construction of the single layer detector was attempted.

Simply speaking, the idea of how to detect second-order singularities is to throw in rotating thin masks in combination with the generalized Laplacian. One inherent trouble of such a detection scheme is manifested immediately by (2.45) and (2.61) : second-order singularities, i.e., curves (or surfaces) across which the derivative of the function undergoes jumps, may be "aliased" by pairs of first order singularities, i.e., "lines."

We cannot end this paragraph on Marr's detector without at least briefly mentioning other notions basic to the considered zero-crossing scheme. Realizing that their detector yields no directional information, the authors of [Ma. Hi.1] introduced the notion of zero-crossing segments and related to it the notion of the slope of zero-crossing. Again, the intuition is that direction of the edge is associated with direction of local zero-crossing segments, which is "proved" to coincide with the direction across which the slope of this zero-crossing is maximal. While the first of these concepts is never made precise, the slope of zero-crossing is just a directional derivative $\frac{\partial}{\partial \ell}(\Delta(f * \varphi))$, where ℓ is any direction of differentiation. In our notation, the assumption above then may be rewritten as

$$\frac{\partial}{\partial \ell}(\Delta(f * \varphi))(\overline{x_0}) \leq \frac{\partial}{\partial \mathbf{n}}(\Delta(f * \varphi))(\overline{x_0}), \tag{2.62}$$

with $\overline{x_0} \in L$ and $\overline{n}$ is normal vector to L. We will show later that when certain conditions are satisfied, (2.62) is valid.

In order to prove this and other useful local extrema estimates in the vicinity of the arbitrary singularity, generalization of *Taylor's Theorem for distributions* is of the utmost importance.

§2.7. *Taylor's Theorem for generalized functions.*

Definition 2.7.1.

Let τ_c be *translation* operation defined for test functions and for generalized functions by

(i) $\tau_C \varphi(\overline{x}) = \varphi(\overline{x} + C)$, $\varphi \in \mathcal{K}_n$, $C \in R^n$ (2.63)

(ii) $\tau_C\, f(\overline{x})$ for generalized function $f(\overline{x}) \in \mathcal{K}_n'$ is given by its action on the arbitrary test function as follows:

$$<\tau_C\, f(\overline{x}), \varphi(\overline{x})> = <f(\overline{x}), \tau_{-C}\, \varphi(\overline{x})>$$
$$= <f(\overline{x}), \varphi(\overline{x} - C)> . \tag{2.64}$$

Just as for test functions, we shall use notation $f(\overline{x} + C) = \tau_C\, f(\overline{x})$, remembering, however, that $f(\overline{x} + C)$ is a *functional*, not a point-value function.

<u>Theorem</u> <u>2.7.1</u>

Let $f(\overline{x}) \in \mathcal{K}_n'$ be an arbitrary generalized function. Then we may write

$$f(\overline{x} + H) = f(\overline{x}) + \sum_{i=1}^{n} H_i \overline{\frac{\partial f}{\partial x_i}} (\overline{x}) + \frac{1}{2} \sum_{i,j=1}^{n} H_i \cdot H_j \overline{\frac{\partial^2 f}{\partial x_i \partial x_j}} (\overline{x}) + \tag{2.65}$$
$$+ \frac{1}{3!} \sum_{i,j,k=1}^{n} H_i H_j H_k \overline{\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}} (\overline{x}) + \ldots + T_n(H, \overline{x}) ,$$

where $\|T_n(H, \overline{x})\| / |H|^n \to 0$ as $|H| \to 0$ in R^n, $H = (H_1, H_2, \ldots, H_n)$ and "bar" stands for our "usual" generalized differentiation.

<u>Proof</u>:

The proof is remarkably analogous to the proof of Taylor's formula for smooth functions. We first introduce a new variable, setting $H = t \cdot h$, where $t \in R$ and $h \in R^n$ is a unit vector. We proceed by the rules of generalized differentiation, translation (2.64), and by the Chain Rule:

$$\frac{d}{dt} \langle f(\overline{x} + t \cdot h), \varphi(\overline{x}) \rangle$$

$$= \frac{d}{dt} \langle f(\overline{x}), \varphi(\overline{x} - t \cdot h) \rangle . \tag{2.66}$$

Differentiating the inner product with respect to the parameter we have

$$= <f(\overline{x}), \frac{d}{dt}\varphi(\overline{x} - t \cdot h)>$$

$$= \left\langle f(\overline{x}), -\sum_{i=1}^{n} \frac{\partial\varphi}{\partial x_i}(\overline{x} - t \cdot h) \cdot h_i \right\rangle ,$$

and transferring differentiation by the integration by parts (2.2)

$$= \left\langle \sum_{i=1}^{n} \frac{\overline{\partial f(\overline{x})}}{\partial x_i} \cdot h_i, \varphi(\overline{x} - t \cdot h) \right\rangle .$$

Applying Definition 2.64 of translation we obtain

$$= \left\langle \sum_{i=1}^{n} \frac{\overline{\partial f(\overline{x} + t \cdot h)}}{\partial x_i} \cdot h_i, \varphi(\overline{x}) \right\rangle .$$

Observing that $\varphi(\overline{x})$ was an arbitrary test function we have

$$\frac{d}{dt} f(\overline{x} + t \cdot h) = \sum_{i=1}^{n} h_i \frac{\overline{\partial f(\overline{x} + t \cdot h)}}{\partial x_i} . \qquad (2.67)$$

On substituting (2.66) into $\frac{d^2}{dt^2}$ we find

$$\frac{d^2}{dt^2} f(\overline{x} + t \cdot h) = \frac{d}{dt} \left(\sum_{i=1}^{n} h_i \cdot \frac{\overline{\partial f(\overline{x} + t \cdot h)}}{\partial x_i} \right) = \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f(\overline{x} + t \cdot h)}{\partial x_i \partial x_j} . \qquad (2.68)$$

Similarly, $\frac{d^3}{dt^3}$ etc. Furthermore,

$$<f(\overline{x} + t \cdot h), \varphi(\overline{x})> = <f(\overline{x}), \varphi(\overline{x} - t \cdot h)>, \qquad (2.69)$$

and by the Taylor expansion of $\varphi(x - t \cdot h)$ as a smooth function of parameter t we obtain

$$= \left\langle f(\overline{x}), \sum_{k=0}^{n} \frac{t^k}{k!} \cdot \frac{\partial^k \varphi(\overline{x} - t \cdot h)}{\partial t^k}\bigg|_{t=0} + T_n(\varphi, t) \right\rangle , \tag{2.70}$$

where the remainder of the Taylor series is given in its differential form

$$T_n(\varphi, t) = \frac{t^{n+1}}{(n + 1)!} \cdot \frac{\partial^{n+1} \varphi(\overline{x} - \theta t \cdot h)}{\partial t^{n+1}}\bigg|_{t=0} \tag{2.71}$$

and $0 < \theta < 1$. Then by the definition of translation, similar to (2.66) we see that (2.69) and (2.70) may continue as

$$= \left\langle \sum_{k=0}^{n} \frac{t^k}{k!} \cdot \frac{\partial^k f(\overline{x} + t \cdot h)}{\partial t^k}\bigg|_{t=0} + T_n(f, t), \varphi(\overline{x}) \right\rangle .$$

Substituting (2.67), (2.68) for $t = 0$ we have

$$= \Big\langle f(\overline{x}) + \sum_{i=1}^{n} \overline{\frac{\partial f(\overline{x})}{\partial x_i}} \cdot t \cdot h_i + \frac{1}{2} \sum_{i,j}^{n} \overline{\frac{\partial^2 f(\overline{x})}{\partial x_i \partial x_j}} \cdot t^2 \cdot h_i \cdot h_j +$$

$$+ \dots + T_n(h \cdot t, \overline{x}), \varphi(\overline{x}) \Big\rangle ,$$

where $|\langle T_n(h \cdot t, \overline{x}), \varphi(\overline{x})\rangle| / t^n \to 0$ as $t \to 0$, due to the remainder's form (2.71). Finally we observe that $\varphi(\overline{x})$ was an arbitrary test function and (2.65) results. Q.E.D.

Remark:

If $f(\overline{x})$ in (2.65) is a surface distribution δ_L and $\varphi(\overline{x})$ is a test function whose support is concentrated in a small neighborhood of $\overline{x_0} \in L$, then (2.65) clearly demonstrates that the action of the singularity on a shifted test function in the vicinity of this singularity depends not only on δ_L, but on *higher generalized derivatives* of δ_L as well.

It is to be shown that these higher-order generalized derivatives carry the information about the *geometry* of singular boundaries of $f(\overline{x})$ so much sought by us.

§2.8 *Calculus of tangential derivatives.*

The question which lies at the heart of any attempt to compute anything associated with possible essential singularities is: what kind of analytical tools should we use to establish a firm computational link between the *outer* world given by the function values $f(\overline{x})$ and the *inner* world of the unknown essential singularities formed by the singular behavior of $f(\overline{x})^2$? In other words, singular boundaries lie hidden somewhere in the extrinsic data obtained by measuring point-values of the function. The only means available to us is manipulating this *extrinsic* data in some clever way so that properties of the *intrinsic* world, originated by the function's singular conduct, may be revealed.

Up to this point we have demonstrated a relation between the first generalized derivative and the direction of the *tangent* vector (or normal) at a point of the singular boundary. And this knowledge was utilized in the construction of the simple edge detecting scheme of "rotating thin masks." In order to establish an intimate relation between the internal geometry of singular boundaries and the "external tools" of higher-order generalized differentiation, the so called Calculus of Tangential Derivatives is applied.

The Calculus of Tangential Derivatives was pioneered by M. Miranda in his 1965 paper [Mir.]. The original purpose of this machinery was an investigation of the properties of minimal surfaces and related quasi-linear elliptic differential equations. Also, it is considered by [Cour. Hilb.] in the discussion of Characteristic Manifolds of partial differential equations of the first order.

In order to facilitate the reader's understanding of its applications to Numerical Analysis of Singularities in the following paragraphs, and because of the scatteredness of this material in existing literature, a brief account of this calculus is given below. The results are due to several authors, as will be indicated in the following exposition.

Definition 2.8.1.

Let E be a bounded open set in R^3 and ∂E its non-empty boundary. The distance function $d(\overline{x})$ is defined by

$$d(\overline{x}) = \text{dist}(\overline{x}, \partial E) , \tag{2.68}$$

where dist $(\overline{x_0}, \partial E)$ is a minimal Euclidean distance between point $\overline{x_0} \in R^n$ and the set ∂E.

Let $\overline{x_0} \in \partial E$ and $\overline{n}(\overline{x_0})$ be a normal vector to ∂E at $\overline{x_0} \in R^n$. We can choose the local coordinate system by having the x_n axis be in the direction of $\overline{n}(\overline{x_0})$ and x_i axis for any $1 \leq i \leq n - 1$ in the direction of principal curvature eigenvectors corresponding to principal curvatures $K_1, K_2, \ldots, K_{n-1}$ of the surface ∂E at $\overline{x_0}$. In fact, if surface ∂E in the neighborhood of $\overline{x_0}$ is formed by the function $x_n = g(x_1, x_2, \ldots, x_{n-1})$, where g is at least twice differential, then the principal curvatures $K_1, K_2, \ldots, K_{n-1}$ are eigenvalues of the Hessian Matrix (D^2g) and our chosen coordinate directions are its eigenvectors. This coordinate system is called *principal*.

If we assume that ∂E is locally described by the equation $G(\overline{x}) = 0$, then with some use of elementary differential geometry (see, e.g., [Car.]) we establish that

$$\overline{n} = \mathbf{grad}\,(G)/|\mathbf{grad}\,G| , \tag{2.69}$$

and that

$$\frac{\partial}{\partial x_i}\left(\frac{\partial G}{\partial x_j}\right) = -K_i \cdot |\mathbf{grad}\,(G)| \cdot \delta_{i,j}$$

$$i, j = 1, 2, \ldots, n-1 \tag{2.70}$$

$$\frac{\partial}{\partial x_i}\left(\frac{\partial G}{\partial x_n}\right) = 0 ,$$

where $\delta_{i,j}$ is a Kronecker symbol.

<u>Definition</u> <u>2.8.2</u>.

Let $K_1, K_2, \ldots, K_{n-1}$ be the principal curvatures of the surface ∂E. Then the mean curvature $H(\overline{x_0})$ and the Gaussian curvature $K(\overline{x_0})$ of $\overline{x_0} \in \partial E$ are defined as follows

$$H(\overline{x_0}) = -\frac{1}{n-1} \cdot \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial G}{\partial x_i}\Big/|\mathbf{grad}\,(G)|\right)(\overline{x_0}) = \frac{\sum_{i=1}^{n-1} K_i}{n-1} \tag{2.71}$$

$$K(\overline{x_0}) = \prod_{i=1}^{n-1} K_i . \tag{2.72}$$

Next we define the notion of tangential derivative.

Definition 2.8.3.

Let $f(\overline{x}) : R^n \to R$ and $\overline{n}$ be a normal vector to the arbitrary smooth surface ∂E in R^n, then the *tangential* gradient vector δf is given by

$$\delta f = \mathbf{grad}(f) - \overline{n} \cdot (\mathbf{grad}(f) \bullet \overline{n}) . \tag{2.73}$$

If we denote the i-th component of the normal vector by n_i, and observe that $\mathbf{grad}(f) \bullet \overline{n} = \frac{\partial f}{\partial \mathbf{n}}$, (2.73) may be rewritten in terms of its components as follows

$$\delta_i f = \frac{\partial f}{\partial x_i} - n_i \cdot \frac{\partial f}{\partial \mathbf{n}}, \text{ for } i = 1, 2, \ldots, n . \tag{2.74}$$

The characteristic property of δf is that it depends *only* on the distribution of the values of $f(\overline{x})$ on the surface ∂E. It is easily shown by observing that if two functions f_1 and f_2 are equal on ∂E, then $g(\overline{x}) = f_1(\overline{x}) - f_2(\overline{x})$ is equal to zero on ∂E, i.e., ∂E is a level curve for $g(\overline{x})$, consequently $\mathbf{grad}\ (g(\overline{x}))_{\overline{x_0}} = \mathbf{grad}(f_1)_{\overline{x_0}} - \mathbf{grad}(f_2)_{\overline{x_0}}$ is normal to ∂E at each $\overline{x_0} \in \partial E$. Hence $\delta_i g = \frac{\partial g}{\partial x_i} - n_i \frac{\partial g}{\partial \mathbf{n}} = \frac{\partial g}{\partial x_i} - n_i \cdot |\mathbf{grad}(g)|$ which yields

$$\delta_i g = \frac{\partial g}{\partial x_i} - \frac{\partial g}{\partial x_i} = 0 .$$

Thus, $\delta f_1 = \delta f_2$ on ∂E and is independent of the values outside ∂E. We may also write formula (2.74) as

$$\delta_i f = D_i f - n_i \cdot \sum_{j=1}^{n} n_j \cdot D_j f , \tag{2.75}$$

with D_i replacing symbol $\left(\frac{\partial}{\partial x_i}\right)$. For the further convenience the tensor summation notation is employed, so that if the same index is repeated twice in the same term, then summation with respect to this index is assumed. For example, (2.75) can be rewritten as

$$\delta_i f = D_i f - n_i \cdot n_j \cdot D_j f \ .$$

It can be shown that the gradient of distance function in (2.68) is equal to the normal vector $\overline{n}$, i.e.,

$$\overline{n} = \mathbf{grad}(d(\overline{x})) \ . \tag{2.76}$$

This implies that $n_i = D_i d$, and

$$\delta_i n_j = D_i n_j - n_i \cdot n_k \cdot D_k n_j \tag{2.77}$$

$$= D_i D_j d - n_i \cdot D_k d \cdot D_k D_j d$$

$$= D_i D_j d - \tfrac{1}{2} n_i \cdot D_j (D_k d)^2 .$$

Inserting the identity $\sum_{k=1}^{n} n_k^2 = D_k d \cdot D_k d_1$ we continue (2.77)

$$= D_i D_j d - \tfrac{1}{2}\, n_i \cdot D_j(|\overline{n}|) = D_i D_j d \ ,$$

and using the commutativity property of mixed derivative obtain

$$= D_j D_i d = \delta_j n_i \ .$$

Thus we have established that

$$\delta_i n_j = \delta_j n_i, \quad n = 1, 2, \ldots, n \,. \tag{2.78}$$

Next, we can apply (2.71) and (2.74) to obtain

$$\delta_i n_i = D_i \left(\frac{D_i G}{|\mathrm{grad}\ G|}\right) - n_i \cdot n_k \cdot D_k\ n_i \,.$$

With $G(\overline{x}) = 0$ being an equation of the surface ∂E, it follows that

$$= -(n - 1) \cdot H - \tfrac{1}{2} \cdot n_k \cdot D_k \sum_{i=1}^{n} n_i^2 = -(n - 1) \cdot H \,,$$

which yields that *mean curvature* H may be computed as

$$H = -\frac{1}{n - 1} \cdot \delta_i\ h_i \,. \tag{2.79}$$

A little calculation from (2.74) shows that

$$\delta_i \delta_k = \delta_k \delta_i + \delta_m (n_i \cdot \delta_k\ n_m - n_k \cdot \delta_i n_m) \cdot \delta_m \,. \tag{2.80}$$

Consequently tangential *mixed* derivatives *do not commute.*

It is useful to prove for later use two more identities. First, by definition (2.74)

$$n_i \cdot \delta_i = n_i \cdot D_i - n_i \cdot n_i \cdot n_k \cdot D_k = n_i \cdot D_i - \sum_{i=1}^{n} n_i^2 \cdot n_k \cdot D_k = 0 \tag{2.81}$$

Similarly the validity is derived for

$$n_i \cdot \delta_k n_i = 0 \ . \tag{2.82}$$

The notation used in the above calculations assumes that operator δ is applied only to the term immediately after it. We may now state the most important property of the tangential derivatives, which will be used over and over again in the following paragraphs. This property will allow us to derive formulas for arbitrary-order generalized derivatives with great ease and in almost routine fashion.

It should be stressed here that the following theorem will allow us to avoid the very unintuitive tensor calculations of [Est. Kan.], who actually pioneered the informal use of what amounts to tangential derivatives in the calculus of generalized functions. But the authors of [Est. Kan.] were apparently unaware of the earlier work of M. Miranda and its subsequent development in the framework of the theory of minimal surfaces. Thus in their approach a whole paper [Est. Kan. 1] is dedicated to the calculation of the higher-order derivatives for simple layers and yet another [Est. Kan.2] to produce *horrendous-looking formulas for just First and Second derivatives of multilayers*; when, in fact, the following theorem will allow us to compute *arbitrary*-order derivatives for *arbitrary* multilayers with great simplicity.

<u>Theorem</u> <u>2.8.1</u> (Morrey [Mor.]).

Let ∂E be a smooth boundary surface in R^{n+1} and $f(\overline{x})$ be some differentiable functions. Then if $H(\overline{x})$ is the mean curvature of the surface

∂E, one has

$$\int_{\partial E} \delta f \cdot d\ell = -n \int_{\partial E} f \cdot H \cdot \bar{n} \cdot d\ell \,, \tag{2.83}$$

where $d\ell$ is a surface element of ∂E.

This crucial formula may be rewritten in the component-wise fashion as

$$\int_{\partial E} \delta_k f \cdot d\ell = -n \cdot \int_{\partial E} f \cdot H \cdot n_i \cdot d\ell \,. \tag{2.84}$$

Now, we are ready to derive the most useful formulas establishing deep *connection between the operation of generalized differentiation and geometry of singular boundaries.*

§2.9 *Exploration of the internal geometry of singularity with higher order generalized derivatives.*

One of the most fundamental concepts of modern differential geometry is that of the *covariant* derivative of the vector filed with respect to the other vector field. The intuition behind this notion is to measure the rate of change of the first field X in the direction provided by the second field Y. It is denoted by the symbol $\nabla_Y X$. Clearly, $\nabla_Y X$ is an extrinsic quantity, for it depends on the space R^n in which fields X and Y are immersed. Thus in case X is just a scalar point function $f : R^n \to R$, $\nabla_Y f$ is simply a directional derivative of f in the direction of the vector field Y at each point in R^n. It is quite natural to define a *surface* *covariant derivative* $\nabla_Y^S X$ which assigns the rate of change of X in the direction Y *as seen from the surface S.* Consequently this derivative is an *intrinsic* measurement which could be studied entirely within surface S without regard to the space R^n in which it is embedded.

It can be shown that $\nabla_Y^S X$ is uniquely defined and is always equal to the tangent component of $\nabla_Y X$ on the surface S. Consequently, it is the projection of the external covariant derivative of the field onto the tangent plane to the surface S.

With the aid of (2.74) it is seen that our *tangential derivative* $\delta_i f$ corresponds to the *intrinsic covariant derivative* $\nabla_{X_i}^{\partial E} f(\bar{x})$, where X_i is a direction field parallel to the axis X_i.

It is well known that covariant derivatives satisfy the linearity and Leibnizian properties. While the linearity is obvious, the rule of tangentially differentiating a product of scalar point-functions on a given surface S would be of great value in our computations. We shall refer to it as the Leibnizian property of tangential derivatives for obvious reasons.

<u>Lemma</u> <u>2.9.1</u>. (Leibnizian property)

Let f and g be differentiable functions, then for any smooth surface ∂E

$$\delta_i(f \cdot g) = f \cdot \delta_i g + g \cdot \delta_i f \,. \qquad (2.85)$$

Proceed by the component-wise formula (2.74)

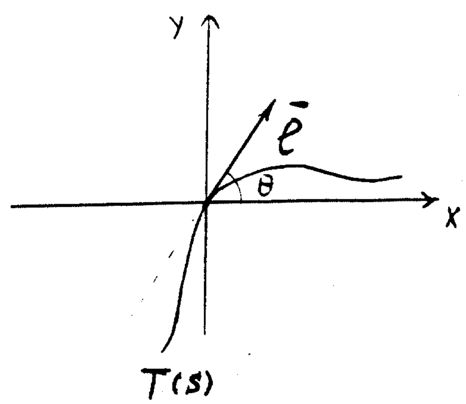
$$\delta_i(f \cdot g) = \frac{\partial}{\partial x_i}(f \cdot g) - n_i \cdot \sum_{k=1}^{n} n_k \cdot \frac{\partial}{\partial x_k}(f \cdot g) \,,$$

assuming again the tensor summation notation

$$= \left(\frac{\partial g}{\partial x_i} - n_i \cdot n_k \cdot \frac{\partial g}{\partial x_k}\right) \cdot f + \left(\frac{\partial f}{\partial x_i} - n_i \cdot n_k \cdot \frac{\partial f}{\partial x_k}\right) \cdot g$$

$$= f \cdot \delta_i g + g \cdot \delta_i f \ .$$

Q.E.D.



(Fig. 2.11)

It is worthwhile to calculate tangential derivatives $\delta_i n_j$ for the case of plane curves, which provide the simplest geometry of singularities. Let the angle $\theta(s)$ be the inclination of the tangent vector to the curve $T(s) = (x(s), y(s))$ on Figure 2.11. Curve T is parametrized with respect to the arc length s. Let $\bar{n}$ and $\bar{\ell}$ be correspondingly unit normal and tangent vectors to T. If such a selection of the parameter is made, we obtain

$$\bar{n} = (n_x, n_y) = (-\dot{y}(s), \dot{x}(s))$$

and

$$\bar{\ell} = (\ell_x, \ell_y) = (\dot{x}(s), \dot{y}(s)).$$

We also have

$$\dot{x}(s) = \cos(\theta(s)), \qquad \dot{y}(s) = \sin(\theta(s)) .$$

It follows that

$$D_x n_x = \frac{\partial n_x}{\partial \ell} \cdot \cos(\theta(s)) + \frac{\partial n_x}{\partial \mathbf{n}} \cdot \cos(n\hat{\ }x) = \dot{n}_x \cdot \dot{x} ,$$

since by (2.76) and identity (2.82)

$$\frac{\partial n_x}{\partial \mathbf{n}} = n_i \cdot D_i\, D_x\, d(\overline{x}) = n_i \cdot D_x D_i d(\overline{x})$$

$$= n_i D_x n_i = 0 .$$

Thus, we have

$$\delta_x n_x = -\ddot{y} \cdot \dot{x} = -\cos^2(\theta(s)) \cdot \dot{\theta}(s) . \tag{2.86}$$

Similarly,

$$\delta_y n_y = \ddot{x}(s) \cdot \sin(\theta(s)) = \ddot{x} \cdot \dot{y} = -\sin^2(\theta(s)) \cdot \dot{\theta}(s) . \tag{2.87}$$

And upon using the symmetry property (2.78)

$$\delta_x n_y = \delta_y n_x = -\ddot{y} \cdot \dot{y} = -\sin(\theta(s)) \cdot \cos(\theta(s)) \cdot \dot{\theta}(s) . \tag{2.88}$$

$\dot{\theta}(s)$ has geometric significance, for it is a *curvature* of T(s), and we write

$$\dot{\theta}(s) = k. \tag{2.89}$$

It is of interest to remark at this point that (2.86)–(2.89) clearly demonstrate the relation between tangential derivatives of the normal vector and the *intrinsic geometry* of curves manifested by its curvature.

In deriving formula (2.45) for the generalized Laplacian we were able to avoid direct computation of $\frac{\bar{\partial}}{\partial x_i}([f] \cdot \delta_{L_k})$ by the use of the second Green's identity. We concluded then that the generalized Laplacian is uniquely decomposable into the sum of regular part and k-multilayers for k = 0, 1. This decomposition has guided our analysis of Marr's zero-crossing schema as well as the line detection algorithm.

There are many reasons why we need to derive formulas of the type of (2.45) in a direct manner. What is meant is the possibility of carrying-out the differentiation process on the terms of FTED decomposition (2.23), obtaining again some unique decomposition into the sum of multilayers, and then if necessary do it over again on the new terms, in the fashion similar to usual differentiation. Of course, the higher the order of differentiation is, the higher the multilayer order will appear in the decomposition. Henceforth formulas for differentiation of arbitrary-order multilayers are required.

The crux of this presentation is application of these formulas to the process of uncovering various computational aspects of possible singularities.

In carrying out this program, a differentiation formula of single layer with arbitrary density is derived.

2.9.1 *Differentiation of single layer.*

Let L be an arbitrary smooth surface in R^n, and $\sigma \cdot \delta_L$ be a single layer on L with smooth density $\sigma : L \to R$. For arbitrary test function $\varphi \in \mathcal{K}_n$ we have by the integration by parts

$$\left\langle \frac{\bar{\partial}}{\partial x_i} \sigma \cdot \delta_L, \varphi \right\rangle = -\left\langle \sigma \cdot \delta_L, \frac{\partial \varphi}{\partial x_i} \right\rangle ;$$

by the definition of the tangential derivative (2.74) we get

$$= -\int_L \sigma \cdot \frac{\partial \varphi}{\partial x_i} \cdot d\ell = -\int_L \sigma \cdot \left[\delta_i \varphi + n_i \cdot \frac{\partial \varphi}{\partial \mathbf{n}}\right] \cdot d\ell . \tag{2.90}$$

Using Leibnizian property (2.85) we see that

$$\sigma \cdot \delta_i \varphi = \delta_i(\sigma \cdot \varphi) - \varphi \cdot \delta_i \sigma . \tag{2.91}$$

We then continue (2.90) as follows:

$$= -\int_L \delta_i(\sigma \cdot \varphi) \cdot d\ell + \int_L \varphi \cdot \delta^i \sigma \cdot d\ell - \int_L n_i \cdot \sigma \cdot \frac{\partial \varphi}{\partial \mathbf{n}} \cdot d\ell .$$

The first integral above can be estimated by application of the Morrey's formula (2.84)

$$= \int_L \left[(n - 1) \cdot H \cdot \sigma \cdot n_i + \delta_i \sigma\right] \cdot \varphi \cdot \bar{d}\ell - \int_L n_i \cdot \sigma \cdot \frac{\partial \varphi}{\partial \mathbf{n}} \cdot d\ell ,$$

which yields *functional* identity

$$\frac{\bar{\partial}}{\partial x_i}(\sigma \cdot \delta_L) = ((n - 1) \cdot H \cdot \sigma \cdot n_i + \delta_i\sigma) \cdot \delta_L + n_i \cdot \sigma \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L \,, \tag{2.92}$$

The most notable feature of formula (2.92) is the appearance of the single layer with the quite remarkable density term. One would expect to get a *double* layer, when the *single* layer is differentiated. Yet unless the surface L is totally trivial, i.e., of zero mean curvature $H = 0$ and the density $\sigma \equiv$ const on L, we see that a single layer emerges as well. And while the density of the double layer carries information only about the component of the normal vector, the density of the single layer explicitly depends on the *curvature* of the surface L and on the *rate of change of* the density on the surface.

By applying (2.92) to FTED (2.23) we can obtain a formula for second-order differentiation. Let $f(\overline{x})$ be as in FTED. To simplify the following calculation we collect all of the singular boundaries in the single symbol L. Then we have

$$\frac{\overline{\overline{\partial f}}}{\partial x_k \cdot \partial x_i} = \frac{\bar{\partial}}{\partial x_k}\left(\frac{\partial f}{\partial x_i} + n_i \cdot [f] \cdot \delta_L\right). \tag{2.93}$$

Now we proceed by applying FTED again to the regular term and (2.92) to the singular component

$$= \frac{\partial f}{\partial x_k \partial x_i} = n_k \cdot \left[\frac{\partial f}{\partial x_i}\right] \cdot \delta_L + \frac{\bar{\partial}}{\partial x_k}(n_i \cdot [f] \cdot \delta_L)\,.$$

The jump of the derivative $\left[\frac{\partial f}{\partial x_i}\right]$ may be easily shown (see [Est. Kan.1]) to be

$$\left[\frac{\partial f}{\partial x_i}\right] = \left[\frac{\partial f}{\partial \mathbf{n}}\right] \cdot n_i + \delta_i[f] \ , \tag{2.94}$$

so that (2.93) with the use of (2.85), and collecting single layer terms may continue as

$$= \frac{\partial f}{\partial x_k \partial x_i} + \delta_L \cdot \left(\left[\frac{\partial f}{\partial \mathbf{n}}\right] \cdot n_i \cdot n_k + n_k \cdot \delta_i[f] + (n-1) \cdot H \cdot n_i \cdot n_k \cdot [f] + \right.$$

$$\left. + n_i \cdot \delta_k[f] + \delta_k\, n_i \cdot [f] \right) + n_i \cdot n_k \cdot [f] \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L \ . \tag{2.95}$$

A number of fruitful observations can be concluded from the quite cumbersome density term of the single layer in (2.95). The most interesting thing about it is the presence of two terms explicitly depending on the mean curvature of the surface L. We also see that the *rate of change of the jump along the singularity contributes to the density* as well.

2.9.2 *Design and analysis of the simple curvature detector.*

By exploiting the occurrence of the *curvature*-dependent terms in the second mixed derivative we can pursue the design and rigorous analysis of the curvature-sensitive singularity detectors. It should be stressed again that the curvature appearing in (2.95) is a *curvature of the singular boundary* and is not a curvature of the surface given by the graph of the image function $f(\overline{x})$. This development is the first totally new practical consequence of the use of generalized derivatives. For it has been observed earlier that the method of "rotating thin masks" in itself is just a very clean theoretical version of the previous edge detecting schemas.

A natural way of motivating the use of curvature-sensitive detection algorithms is to look at what kind of computational information is conveyed by curvatures and how it can be utilized for approximation of the singular surfaces.

Let R(t) be a vector parametric form of the singular boundary L. Then utilizing Taylor expansion of R(t) we have

$$\Delta R(t) = R(t + \Delta t) - R(t) \tag{2.96}$$

$$= R'(t) \cdot \Delta t + \tfrac{1}{2} \cdot R''(t) \cdot \Delta t^2 + \tfrac{1}{6} \cdot R'''(t) \cdot \Delta t^3 + \dots .$$

Expansion (2.96) demonstrates that up to terms of the order $O(\Delta t)$, curve L coincides with its tangent line in the neighborhood of each point, i.e.,

$$\Delta R(t_0) = R'(t_0) \cdot \Delta t + O(\Delta t) .$$

We also see that up to $O(\Delta t^2)$, R(t) lies in the plane determined by the vectors R'(t) and R''(t), known as an *osculating plane* of curve L. Therefore, up to $O(\Delta t^2)$, curve L is locally a *parabola* in the coordinate system given by the basis $(R'(t), \frac{1}{2}R''(t))$. Indeed, its parametric form is then

$$x = \Delta t \text{ and } y = \Delta t^2.$$

Analogously we conclude that up to $O(\Delta t^3)$, L resides in the 3-dimensional subspace given by the basis $(R'(t), \frac{1}{2} R''(t), \frac{1}{6} R'''(t))$. In this subspace R(t) is locally described by the third-order curve.

Assuming natural parametrization of the curve L and making use of the Frenet formulae, we can rewrite by identity (2.96) as follows,

$$\Delta R(s) = \ell \cdot \Delta s + \tfrac{1}{2} k \cdot \overline{n} \cdot \Delta s^2 + \tfrac{1}{6} \cdot (-k^2 \cdot \overline{\ell} + k' \cdot \overline{n} + k \cdot \tau \cdot \overline{b}) \cdot \Delta s^3 + 0(\Delta x^3), \quad (2.97)$$

with $\{\overline{\ell}, \overline{n}, \overline{b}\}$ forming a Frenet trihedron and where k is a *curvature* of L and τ is its torsion. The above consideration implies that tangent, curvature, torsion and "higher curvatures" provide coordinates in the local basis associated with each point on the curve and given by

$$(R'(t), R''(t), \ldots, R^{(n)}(t)).$$

In particular, for planar curves, the curvature corresponds to the *second dimension* given by the normal vector, when the first dimension is assigned to the tangent vector. Thus it is curvature k, which literally "curves" our curve L from the would-be straight line in the direction of the tangent. Similarly, the torsion provides the *third* dimension, if L is a space curve.

This discussion on the geometrical significance of the first and higher curvatures furnishes the interpretation of all previously derived detection schemas as essentially one-dimensional. The precise meaning of the above statement follows from the Fundamental Theorem of Edge Detection and subsequent analysis in §2.4 and §2.5.

It has been firmly established there that regardless of the different appearances of various more or less "successful" edge detectors of today, they all work on the basis of the same fundamental principle, expressed by (2.23) of FTED. But singular contribution in (2.23) can only yield the estimate of the jump in the direction of the *normal* to the singular boundary, which we are fortunate to discover only under the best of circumstances analyzed in §2.4. Consequently, it is this direction of the *normal* or equivalently the direction of *tangent* which these detection schemas are going after. Thus in light of the above discussion on the curvatures and dimensionality of the curve (surface) all these methods, including our "rotating masks," are essentially uni-dimensional detectors. Now that we know the computational importance of the curvature we see that even the best schemas of today are bound to fail on any nontrivial, highly curved singular boundaries. For all these schemas are fundamentally blind to the "second dimension" of the planar curve boundaries, i.e., to curvatures, or to higher dimensions such as torsion when the problem of the singularity detection is considered in R^3 or higher R^n. As an application of (2.95) we can construct and analyze a "simple linear curvature detector" for planar boundaries. To obtain such a detector we merely replace the *first* generalized derivative in the familiar symbolic expression of the first-order "thin masks" detector $\left\langle\frac{\overline{\partial} f(\overline{x})}{\partial \ell}, \varphi_\alpha\right\rangle$ by the *second*-order generalized derivative $\left\langle\frac{\overline{\partial} f}{\partial \ell \cdot \partial n}, \varphi_\alpha\right\rangle$ where ℓ and n are two mutually perpendicular directions of the differentiation in R^2, $\varphi_\alpha \in \mathcal{K}_2$ and $f : R^2 \to R$. Let us first assume for simplicity that φ_α is a circularly symmetric test function (2.1.0). We can further simplify assumptions about function $f(\overline{x})$ by supposing that the rate of change of jump [f] within supp(φ) and the regular contribution are both negligible, and the jump of the normal derivative $\left[\frac{\partial f}{\partial n}\right]$ is very small as well. We choose coordinate system so that the pair of orts $(\overline{x}, \overline{y})$ corresponds to the directions $(\overline{\ell}, \overline{n})$ of the detector. We shall now expand the detector under consideration according to (2.95). By the integration by parts

$$\left\langle \frac{\overline{\partial f}}{\partial x \partial y}, \varphi \right\rangle = \left\langle f, \frac{\partial \varphi}{\partial x \partial y} \right\rangle , \tag{2.98}$$

and (2.98) reflects the actual way to compute the detector's action on the test function and $f(\overline{x})$. On the other hand, by (2.95) and simplifying the assumption about $f(\overline{x})$ we have

$$= \left\langle (k \cdot n_x \cdot n_y \cdot [f] + \delta_x n_y \cdot [f]) \cdot \delta_L, \varphi \right\rangle +$$

$$+ \left\langle n_x \cdot n_y \cdot [f] \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L, \varphi \right\rangle,$$

where $\overline{n} = (n_x, n_y)$ is outer normal to L and k is the variable curvature of L. We observe that the second-layer term behaves roughly as a first-order detector since its density depends only on the size of the jump [f] and directions of differentiation. But the first term exhibits more interesting behavior. If we assume also that the approximate zero-crossing condition is actually met by the singularity under consideration, then the second term vanishes and upon application of identities (2.88) and (2.89) we have

$$= \langle -2\sin(\theta(s)) \cdot \cos(\theta(s)) \cdot k \cdot [f] \cdot \delta_L, \varphi \rangle$$

$$= -\int_L k \cdot \sin(2\theta) \cdot [f] \cdot \varphi \cdot d\ell \, . \tag{2.99}$$

(See Fig. 2.11). This detector will assume maxima in the direction $\pi/4$ from the direction of the corresponding first-order detector. Thus taking a linear combination of both

$$\alpha_1 \cdot \left\langle \frac{\overline{\partial f}}{\partial \ell}, \varphi_1 \right\rangle + \alpha_2 \cdot \left\langle \frac{\overline{\partial f}}{\partial \ell \cdot \partial n}, \varphi_2 \right\rangle , \tag{2.100}$$

with possibly differently shaped masks, we may obtain some interaction of the tangential direction with the curvature direction, which gives us a fairly general way to approximate singular boundaries by the sequence of linear chord segments. Computational implementation of this detector and a few computed examples are discussed in the last paragraph of this chapter.

A quick look at the expression for the generalized Laplacian (2.45) will reveal that $\overline{\Delta}f(\overline{x})$ does not contain any curvature-dependent term, which is surprising in view of the fact that the generalized Laplacian is a combination of the *second-order* generalized derivatives, each of which, accordingly to (2.95), contains a nonzero, curvature-dependent, simple-layer term. Of course, we have derived (2.45) in a manner which did not involve actual computation of the generalized derivatives, yet the result should be the same no matter how it has been obtained. A simple explicit computation of $\frac{\overline{\partial f}}{\partial x^2} + \frac{\overline{\partial f}}{\partial y^2}$ via (2.95) and (2.86), (2.87) alleviates our concern by showing how curvature-dependent terms annhilate each other in the generalized Laplacian $\overline{\Delta}f(\overline{x})$.

Of much greater interest is to obtain a detector whose curvature contribution is analogous to the generalized gradient magnitude detector (2.39), i.e., is independent from the directions of differentiation. In somewhat mathematically loose fashion we should see that a known *Beaudet corner detector* [Bea.] replaced by its *generalized* version, in fact, is such a rotation-invariant *curvature detector*. Its sensitivity to discontinuities of the boundary will be also meticulously analyzed in the next paragraph.

In what follows we will need rules for *multiplying* simple layers defined on the same surfaces. Disappointingly, the modern theory of generalized function does not possess such machinery as yet, except for very special cases. Having historical precedent in the brave way in which P. Dirac operated with the very first singular distributions 60 years ago, distributions will be multiplied here in rather cavalier fashion, based only on crude intuition borrowed from the sequential approach to the Irregular operations on distributions by J. Mikusinski [Mik.].

Again, with the previous simplifying assumptions by (2.98) and expressions for $\delta_i h_j$ in (2.86)–(2.89) we have

$$\overline{\frac{\partial^2 f}{\partial x^2}} = -k \cdot [f] \cdot \cos(2\theta) \cdot \delta_L + \sin^2(\theta) \cdot \frac{\partial}{\partial n}\, \delta_L \qquad (2.101)$$

$$\overline{\frac{\partial^2 f}{\partial y^2}} = k \cdot [f] \cdot \cos(2\theta) \cdot \delta_L + \cos^2(\theta) \cdot \frac{\partial}{\partial n}\, \delta_L\ . \qquad (2.102)$$

Furthermore, exploiting the fact that reaction of the test function to the second-order layer is weak, or zero on the boundary (under our assumptions), we "formally" multiply simple layers in (2.101) and (2.102) and square simple layer in (2.99) to obtain

$$B(f) = \left(\overline{\frac{\partial^2 f}{\partial x \partial y}}\right)^2 - \overline{\frac{\partial^2 f}{\partial x^2}} \cdot \overline{\frac{\partial^2 f}{\partial y^2}} = k \cdot [f] \cdot \delta_L\ , \qquad (2.103)$$

where we multiplied simple layers on the surface L by simply multiplying the density terms. It should also be noticed that (2.103) is only an approximate equality, since the second-layer terms are actually contributing to the result of B(f).

Thus we see that the so called Beaudet's "corner" detector [Bea.] is nothing else but a rotation-invariant curvature detector. This detector will assume maxima, when the segment of the high curvature lies within the support of the detecting mask. Similar to the previous cases, the generalized differentiation is transferred via (2.2) as follows:

$$<B(f), \varphi> \approx <f, \varphi_{xy}>^2 - <f, \varphi_{xx}> \cdot <f, \varphi_{yy}> , \tag{2.104}$$

where the subscript is a partial differentiation symbol. The reader will notice that expression (2.104) is not the inner product of B(f) in (2.103) with test function $\varphi(\overline{x})$ but is an approximation by the componentwise application of the multiples in (2.103). We shall return to the curvature-sensitive functionals in the analysis of singular boundaries.

Another interesting application of the identity (2.92) of the derivative of a single layer is analysis of the *non-maximum suppression* scheme of Canny (see p. 43 [Can.]). In this scheme, the edge at $(\overline{x}_0)$ is determined upon satisfying the following condition:

$$\left(\frac{\partial^2}{\partial n^2} \; G * I\right)(\overline{x}_0) = 0 , \tag{2.105}$$

where I is an image and G is a smoothing symmetric Gaussian, and $\frac{\partial^2}{\partial n^2}$ means differentiation in the direction provided by the gradient of the function it is applied to. The rationale behind this detection algorithm is similar to the Marr's zero-crossing scheme, though this time maxima or minima of the *first directed* derivative are located in the direction of the gradient, which is assumed to correspond to the direction of the prospected edge.

This last assumption has been shown to be a questionable one, particularly in the neighborhood of nontrivial singular boundaries (see §2.4). Now we shall see directly how curvature of the boundary affects condition

(2.105) of the $\frac{\partial^2}{\partial \mathbf{n}^2}$ operator. First observe that

$$\frac{\partial}{\partial \mathbf{n}}(f * \varphi) = \sum_i n_i \cdot \frac{\partial}{\partial x_i}(f * \varphi) , \qquad (2.106)$$

where h_i are components of the gradient. Using tensor notation and assuming that gradient vector is faithful to the direction of the normal of the singularity of $f(\overline{x})$ we have

$$= \left\{n_i \cdot \frac{\partial f}{\partial x_i} + n_i^2 \cdot [f] \cdot \delta_L\right\} * \varphi .$$

Observe that $\sum_i n_i^2 = 1$. Combining this result with the formula for the derivative of a single layer (2.92) we have

$$\left\langle \frac{\overline{\partial^2}}{\partial \mathbf{n}^2} f, \varphi \right\rangle = \left\langle \frac{\overline{\partial}}{\partial \mathbf{n}} \left\{ n_i \frac{\overline{\partial f}}{\partial x_i} \right\} + \frac{\overline{\partial}}{\partial \mathbf{n}} ([f] \cdot \delta_L), \varphi \right\rangle . \qquad (2.107)$$

If we neglect the regular contributions and jumps of the higher derivatives of f across L, application of (2.106) leads to

$$= \left\langle \frac{\overline{\partial}}{\partial x_i} ([f] \cdot \delta_L), n_i \cdot \varphi \right\rangle$$

$$= \left\langle [f] \cdot \delta_L, \frac{\partial}{\partial x_i} (n_i \cdot \varphi) \right\rangle .$$

Substituting for the $\frac{\partial}{\partial x_i}$ derivative its expression in terms of the *tangential* and normal components as in (2.74), we have

$$= -\left\langle [f] \cdot \delta_L, \quad \delta_i(n_i \cdot \varphi) + n_i^2 \frac{\partial \varphi}{\partial \mathbf{n}} \right\rangle .$$

Application of the Leibnizian property of tangential derivatives and formula (2.79) for the mean curvature yields

$$= -\left\langle [f] \cdot \delta_L, \quad \delta_i n_i \cdot \varphi + n_i \cdot \delta_i \cdot \varphi + \frac{\partial \varphi}{\partial \mathbf{n}} \right\rangle .$$

Using identity (2.81) we obtain

$$= -\left\langle [f] \cdot \delta_L, \quad -(n - 1) \cdot H \cdot \varphi + \frac{\partial \varphi}{\partial \mathbf{n}} \right\rangle .$$

Substituting expression for the double layer (2.14) we get

$$= \left\langle (n - 1) \cdot H \cdot [f] \cdot \delta_L + \frac{\partial}{\partial \mathbf{n}} \delta_L, \varphi \right\rangle .$$

Finally, if we account for the contributions of the regular term and jump of the derivative, the expression for the nonmaximum suppression becomes

$$S(f) = \frac{\overline{\partial f}}{\partial \mathbf{n}^2} = \frac{\partial^2 f}{\partial \mathbf{n}^2} + \left(\left[\frac{\partial f}{\partial \mathbf{n}} \right] + (n - 1)\, H \cdot [f] \right) \cdot \delta_L + [f] \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L \, , \qquad (2.108)$$

where H is mean curvature of the singularity surface L. It is clear from (2.108) that if curvature H is zero, the nonmaximum suppression is almost equivalent to the zero-crossings of the Laplacian given by zeros of $\overline{\Delta} f$ in (2.45), except for directed regular normal derivative $\frac{\partial^2 f}{\partial \mathbf{n}^2}$ instead of the regular Laplacian Δf. This difference will be negligible when a narrow mask is used in (2.107).

In the same vein we should note that the zero-crossing condition (2.107) is severely perturbed by the nonzero curvature contribution in

(2.108), making this scheme extremely inaccurate in the vicinity of the curved edges.

Further analysis of higher curvatures of the boundaries as well as analysis of the generalized functions' extremal behavior on the singularity boundaries can be only achieved with methods operating on the *third-order* generalized derivatives. To evaluate these higher-order distributional derivatives, we need to be able to differentiate *double layers*.

2.9.3. *Differentiation of the double layer.*

Let L be an arbitrary smooth surface in R^n and $\sigma : L \to R$ be a smooth density function of the *double* layer on L.

Our interest is in deriving formula for the generalized derivative of the double layer $\left[\sigma \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L\right]$. We consider

$$\left\langle \frac{\overline{\partial}}{\partial x_i} \left[\sigma \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L\right], \varphi \right\rangle = -\left\langle \sigma \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L, \frac{\partial \varphi}{\partial x_i} \right\rangle . \tag{2.109}$$

Then by definition (2.14) of the double layer we have in tensor summation

$$= \left\langle \sigma \cdot \delta_L, n_k \cdot \frac{\partial^2 \varphi}{\partial x_k \partial x_i} \right\rangle ,$$

or changing order of the differentiation we have

$$= \left\langle \sigma \cdot \delta_L, n_k \cdot \frac{\partial}{\partial x_i} \left[\frac{\partial \varphi}{\partial x_k}\right] \right\rangle ,$$

by (2.74) and the Leibnizian property

$$= \left\langle \sigma \cdot \delta_L, \frac{\partial}{\partial x_i} \left[n_k \cdot \frac{\partial \varphi}{\partial x_k} \right] - \delta_i n_k \cdot \frac{\partial \varphi}{\partial x_k} \right\rangle ,$$

by the definition of the normal derivative we write

$$= \left\langle \sigma \cdot \delta_L, \frac{\partial}{\partial x_i} \; \frac{\partial \varphi}{\partial \mathbf{n}} \right] - \delta_i n_k \cdot \frac{\partial \varphi}{\partial x_k} \right\rangle . \tag{2.110}$$

Separate (2.110) into the difference of two terms

$$= A - B ,$$

where $A = \left\langle \sigma \cdot \delta_L, \frac{\partial}{\partial x_i} \; \frac{\partial \varphi}{\partial \mathbf{n}} \right\rangle$ and $B = \left\langle \sigma \cdot \delta_L, \delta_i n_k \cdot \frac{\partial \varphi}{\partial x_k} \right\rangle$. Carrying out density term to the right and applying Leibnizian property and (2.74) we get

$$A = \left\langle \delta_L, \sigma \cdot \left[\delta_i \frac{\partial \varphi}{\partial \mathbf{n}} + n_i \cdot \frac{\partial^2 \varphi}{\partial \mathbf{n}^2} \right] \right\rangle , \tag{2.111}$$

by the Leibnizian property

$$= \left\langle \delta_L, \delta_i \left[\sigma \cdot \frac{\partial \varphi}{\partial \mathbf{n}} \right] - \delta_i \sigma \cdot \frac{\partial \varphi}{\partial \mathbf{n}} + \sigma \cdot n_i \cdot \frac{\partial^2 \varphi}{\partial \mathbf{n}^2} \right\rangle .$$

We may now apply Morrey's theorem (2.83)

$$= \left\langle \delta_L, -(n-1) \cdot \sigma \cdot n_i \cdot H \cdot \frac{\partial \varphi}{\partial \mathbf{n}} - \delta_i \sigma \cdot \frac{\partial \varphi}{\partial \mathbf{n}} + \sigma \cdot n_i \cdot \frac{\partial^2 \varphi}{\partial \mathbf{n}^2} \right\rangle ,$$

transferring generalized differentiation back to the left as in (2.15) we get

$$= \left\langle \left[(n-1) \cdot H \cdot \sigma \cdot n_i + \delta_i \sigma \right] \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L + \sigma \cdot n_i \cdot \frac{\partial^2}{\partial \mathbf{n}^2} \delta_L, \varphi \right\rangle .$$

Similarly we have

$$B = \left\langle \delta_L, \sigma \cdot \delta_i n_k \cdot \left[\delta_k \varphi - n_k \cdot \frac{\partial \varphi}{\partial \mathbf{n}}\right]\right\rangle \tag{2.112}$$

$$= \left\langle \delta_L, \sigma \cdot \delta_i n_k \cdot \delta_k \varphi - \sigma \cdot \delta_i n_k \cdot n_k \frac{\partial \varphi}{\partial \mathbf{n}}\right\rangle .$$

Using identity (2.82) and the Leibnizian property we write

$$= \left\langle \delta_L, \delta_k(\sigma \cdot \delta_i n_k \cdot \varphi) - \varphi \cdot \delta_k(\sigma \cdot \delta_i n_k)\right\rangle .$$

Application of the Morrey's theorem (2.83) implies

$$= \left\langle \delta_L, (n - 1) \cdot H \cdot \sigma \cdot n_k \cdot \delta_i n_k \cdot \varphi - \delta_k(\sigma \cdot \delta_i n_k) \cdot \varphi\right\rangle .$$

Upon application of the identity (2.82) we write

$$= -\left\langle (\sigma \cdot \delta_k \delta_i n_k + \delta_k \sigma \cdot \delta_i n_k) \cdot \delta_L, \varphi\right\rangle .$$

Finally, collecting (2.111) and (2.112) we obtain a formula for the derivative of the double layer:

$$\frac{\overline{\partial}}{\partial x_i}\left(\sigma \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L\right) = (\sigma \cdot \delta_k \delta_i n_k + \delta_k \sigma \cdot \delta_i n_k) \cdot \delta_L +$$
$$+ ((n - 1) \cdot H \cdot \sigma \cdot n_i + \delta_i \sigma) \cdot \frac{\partial}{\partial \mathbf{n}} \delta_L + \sigma \cdot n_i \cdot \frac{\partial^2}{\partial \mathbf{n}^2} \delta_L . \tag{2.113}$$

Formula (2.113) demonstrates the appearance of the higher curvature term $\delta_k \delta_i n_k$, which in the case of the three-dimensional surface L will carry information about the *torsion* coefficient of L and thus could be utilized in the *detection of singularity surfaces with torsion.*

In order to obtain formulae for the third-order differentiation of the nonsmooth function $f(\overline{x})$, we combine (2.113) and (2.92) to get

$$\frac{\partial^2}{\partial^2 x_i}(\sigma \cdot \delta_L) = \frac{\partial}{\partial x_i}(K_1 \cdot \delta_L + K_2 \cdot \frac{\partial}{\partial n}\delta_L), \tag{2.114}$$

where $K_1 = (\delta_i \sigma + H \cdot \sigma \cdot n_i)$, $H_1 = (n - 1) \cdot H$ and $K_2 = n_i \cdot \sigma$, then combine (2.114)

$$= (\delta_i K_1 + H_1 \cdot K_1 \cdot n_1) \cdot \delta_L + n_i \cdot K_1 \cdot \frac{\partial}{\partial n}\delta_L + \frac{\partial}{\partial x_i}\left[K_2 \cdot \frac{\partial}{\partial n}\delta_L\right]$$

$$= (\delta_i K_1 + H_1 \cdot K_1 \cdot n_1 + \delta_m K_2 \cdot \delta_i n_m + K_2 \cdot \delta_m \delta_i n_m) \cdot \delta_L +$$

$$+ (n_i \cdot K_1 + n_i \cdot H_1 \cdot K_2 + \delta_i K_2) \cdot \frac{\partial}{\partial n}\delta_L + \sigma \cdot \frac{\partial^2}{\partial n^2}\delta_L, \tag{2.115}$$

which can be further unfolded substituting expressions for the coefficients K_1 and K_2 and use of the Leibnizian property and simplifying identities

$$n_m \delta_m = 0, \quad n_m \delta_i n_m = 0 \text{ and } H = \delta_m n_m .$$

Corollary (2.61) gave an answer to the question of why Marr's zero crossing of Laplacian scheme works for simple singularity boundaries. It has been concluded there that even the elongated zero-crossing detector yields no directional information about the normal of the boundary. The notion of the *direction of the zero-crossing* is used in [Ma. Hi.1] to alleviate the above problem. The resulting criterion of the maximal slope of zero-crossing has been rewritten as inequality (2.62), which we have promised to prove. The next lemma demonstrates that only under very simplifying assumptions will Marr's criteria work.

Lemma 2.9.2. (On direction of zero-crossing).

Let L be a circular essential boundary, and supp($\varphi(\bar{x})$) small enough to assume $[f]$=const, $\left[\frac{\partial f}{\partial n}\right]$=const and $\left[\frac{\partial^2 f}{\partial n^2}\right]$=const, then **grad** ($\bar{\Delta}$ u $*$ φ) is normal to the essential boundary L.

Proof.

Taking into account our assumptions, we write out the expression for the singular part of $\frac{\partial}{\partial x_i}$ ($\bar{\Delta}$f) based on the formulae (2.115), (2.95), (2.92):

$$\frac{\partial}{\partial x_i}(\bar{\Delta} f) = \left(\left[\frac{\partial^2 f}{\partial n^2}\right] - C^2 \cdot [f]\right) \cdot n_i \cdot \delta_L + \left(\left[\frac{\partial f}{\partial n}\right] + H_1 \cdot [f]\right) \cdot n_i \cdot \frac{\partial}{\partial n}\delta_L +$$

$$+ [f] \cdot n_i \cdot \frac{\partial^2}{\partial n^2}\delta_L .$$

where $C^2 = \delta_K n_m \cdot \delta_K n_m$ is the sum of the squares of the principal curvatures of the L. The symmetry of the $\varphi(\bar{x})$ and circular assumption about L imply that $\frac{\partial^K \varphi}{\partial x^K}$, k = 0, 1, ... is symmetric across the point P on Figure 2.7. Hence substituting (2.116) into (2.33) and proceeding according to the subsequent proof in Lemma 2.2.3, we arrive at the conclusion that $\nabla(\bar{\Delta}f * \varphi)$ *is normal to* L. Q.E.D.

§2.10. ***Computing singular singularities.***

All differentiation formulas of generalized calculus considered so far can be characterized as calculus of "smooth" singularities. It sounds almost contradictory. Indeed, how can a singularity be smooth? If we look back to the definition of k-th multilayer (2.15), we may ask a question of what happens if the *density term of the multilayer is not a smooth function* or if *the surface* L *on which multilayer is concentrated is not a smooth surface*.

For example, the last possibility is manifested by the *intersecting surfaces* of discontinuity (or generally, surfaces of nonsmoothness) of the regular function $f(\overline{x})$ in (2.23) of FTED. It may also depend on the viewpoint of the observer, whether to consider the density term to be discontinuous or instead to subdivide the boundaries into nonsmooth pieces with smooth corresponding density terms. This possibility is given by the intensity function of the two-dimensional checker board.

Breaks in the smoothness of singular boundaries have great practical implications in various applications of image analysis. It is common in the literature on Shape Analysis (see [Pavl.] to refer to certain "singular" features of geometric shapes such as "corners" or "smooth joins." For the purposes of identification and pattern recognition, various singular sites of the outlines may be first determined and then the shape can be approximated by passing a spline through these points. This approach tries to eliminate the complete reliance on the output of the usual edge detector; for as we know by now, one cannot trust in its measurements with today's state of the art in edge detection. Besides, the noise problem further complicates the shape recognition. However, if some prominent boundary feature is present, additional possibilities arise.

In [Asa. Bra.] Marr's zero-crossing detection scheme has been extended to the task of finding significant curvature changes along jump-singular boundaries in images. First, edge is detected by the Canny's scheme [Can.] which is basically similar to our narrow masks approach without curvature correction terms and without explicit analytical form for the detector's test function it is interesting to note ([Can.] derives his 1-dimensional mask from the stochastic optimization process). Then direction of the normal is estimated as a function of the natural parameter, i.e., arc length. The resulting 1-dimensional function is probed via zero-crossing method for the presence of various combinations of two basic types of nonsmooth behavior of the edge. A "*corner*" is a point where *direction of the tangent* (or equivalently the normal) changes in discontinuous fashion. And *smooth join*—a point across which *curvature* (i.e., derivative of the tangent vector function) undergoes a jump. These two types of behavior could be then compounded into the "end," "crank," "bump," etc. Finally, points of the significant response are matched across several scales and curvature primitives are identified. As we noted already, important applications vary from object recognition to the industrial quality control automated visual inspection. A further advantage of the nonsmooth singular boundaries is taken in the so called structure from motion approach. It is observed there that tracking "interesting points" from frame to frame, the correspondence problem and computation of the optical flow may be achieved. For example, time varying corners [Sha. Jai.] allow us to circumvent difficulties associated with the *aperture problem* of motion detection, which manifests itself in the impossibility to compute tangential component of optical velocity on the smooth boundaries.

I would like to claim that these applications of singular boundaries are just the tip of the iceberg. In particular, numerical analysis of the arbitrary time-space developing singularities, which are submanifolds of the singularity boundaries (3-dimensional surfaces) is of great interest if one is to succeed in the meaningful information extraction from the *time varying imagery*.

Unlike the case of edge detection, only handful of the ad hoc methods exist today to solve the simplest of this circle of problems. Namely, aside from algorithms built on top of the usual edge detection as in the above [Asa. Bra.], few corner detectors were proposed. Moreover, [Sha. Jai.] have shown that all of them are essentially equivalent to the already encountered here Beaudet corner detector [Bea.]. However, it has been demonstrated in (2.103) that *generalized* Beaudet scheme is just a curvature sensitive functional, i.e., *a curvature detector*.

Consequently, we may ask a question: why does a curvature detector also react strongly with corners, which correspond to the points where boundary normal undergoes sudden jump? Having answered this question, we will be able to indicate a more general approach to the *direct detection* of the *singular submanifolds of the singularity surfaces*. By a direct detection we mean a method based on the direct interaction with the image intensity function, without intermediate use of the edge detection scheme, for determination of the curvature behavior cannot rely on the accuracy of today's uni-dimensional detecting algorithms. Besides, direct methods are much less computationally expensive.

Just as in the case of singularity boundaries, the difficulties can be surmounted by associating certain distributions with *singularities of the boundaries* in such a way that characteristic singular contributions may be calculated.

Let us first consider an elementary example of the nonsmooth boundary given by the Heaviside function.

$$H(x, y) = \begin{cases} 1, \ x \geq 0, \ y \geq 0 \\ 0, \ \text{otherwise} \end{cases}$$

Then we see that

$$\left\langle \frac{\partial}{\partial x \partial y} (H(x, y)), \varphi \right\rangle = -\left\langle \delta_{L_2}, \frac{\partial \varphi}{\partial y} \right\rangle$$

where L_2 is positive y-axis, and by the fundamental theorem of calculus and boundness of the supp(φ) we have

$$= \varphi(0,0)$$

Hence we can conclude that

$$\frac{\overline{\partial H}}{\partial x \partial y} = \delta_{(0,\,0)}\,,$$

i.e., a *two-dimensional point delta function.*

In fact we can see that if ℓ is direction tangential to the two-dimensional boundary L with end points P_1 and P_2 and φ is a test function then generalized derivative in the tangential direction of the line-delta function

$$\left\langle \frac{\partial}{\partial \ell}\, \delta_L, \varphi \right\rangle = -\int_L \frac{\partial \varphi}{\partial \ell} \cdot d\ell \tag{2.117}$$

and if L is parametrically given by the $L(t) = (x(t), y(t))$, then

$$\frac{\partial \varphi}{\partial \ell} = \frac{1}{|L'(t)|} \cdot \frac{\partial \varphi}{\partial t} \text{ and } d\ell = |L'| \cdot dt,$$

consequently we may continue (2.117)

$$= -\int_{t_0}^{t_1} \frac{d}{dt}\,(\varphi(t)) \cdot dt = -\Big|_{t_0}^{t_1} \varphi(t)$$

Hence if P_1 and P_2 are the end points of the singular boundary L the following is true

$$\frac{\partial}{\partial \ell}\, \delta_L = \delta_{P_1} - \delta_{P_2} \tag{2.118}$$

where δ_{P_i} are *point two-dimensional distributions.*

Already we have touched on the subject of detection of line and surface distributions. The point distributions are no different. To detect them we can simply use small circular masks, whose support area is vanishing, yet the contributions of the point singularities as in (2.118) are preserved.

If we try to derive (2.118) through the use of Miranda's tangential derivatives the essential new construct would have to be added to it.

In order to eliminate difficulties presented by the discontinuous functions, the regular calculus had to be extended to the calculus of generalized functions. Similarly, to overcome limitations of the calculus of tangential derivatives when boundaries *with singularities* are encountered, we can go one step further and introduce *generalized tangential derivatives*. To my knowledge, there has been no development of this kind in mathematical literature with the exception of a few brief remarks on p. 483 of [Est. Kan.2].

If the *principal* coordinate system of §2.8 is introduced *locally* near the n-dimensional surface of singularity L, then for functions given *on this surface* the divergence theorem is still valid when applied with respect to the differentiation in the tangent plane, i.e., covariant derivative. Then it is not surprising that formulas similar to (2.23) of FTED is obtainable for discontinuous functions given on surface L. Following [Est. Kan.2] we then write

$$\overline{\delta_i f} = \delta_i g + m_i \cdot [g] \cdot \delta_S \qquad (2.119)$$

where "bar" stands for the generalized tangential derivative of $g : L \to R$, $[g]$ is a jump function of g across submanifold $S \subset L$, and δ_S is our usual delta function, concentrated on submanifold S, with m_i being a component of the *tangential* normal to S in the direction X_i.

Substituting (2.119) into the differentiation formula (2.92) for smooth simple layer we *formally* obtain

$$\frac{\overline{\partial}}{\partial x_i}(g \cdot \delta_L) = ((n-1) \cdot H \cdot g \cdot n_i + \delta_i g + m_i \cdot [g] \cdot \delta_S) \cdot \delta_L + \qquad (2.120)$$
$$+ n_i \cdot g \cdot \frac{\partial}{\partial n} \delta_L$$

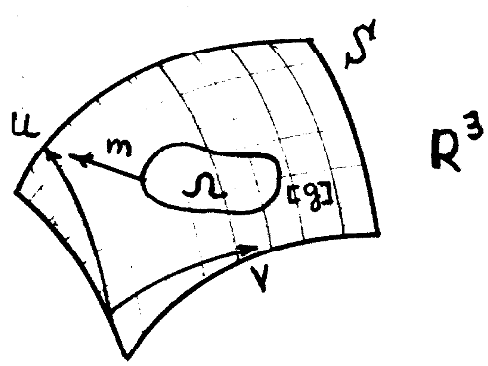
Consequently, we encounter here the *product* of the two simple layers, namely $m_i \cdot [g] \cdot \delta_S \cdot \delta_L$. But the formal expression (2.92) has a precise computational meaning as a functional which acts according to rules (2.15) on the space of test functions. Yet formal expression $\sigma \cdot \delta_S \cdot \delta_L$ in (2.120) has never been given rigorous meaning.

It follows from the tangential FTED (2.119) that *the product of the two simple layers is a simple layer concentrated on their intersection*. Since in our case $S \subset L$, it follows

$$\sigma \cdot \delta_S \cdot \delta_L = \sigma \cdot \delta_S \tag{2.121}$$

Observe that formula (2.118) can now be derived in the direct fashion by noticing that the density term of δ_L in (2.117) is a discontinuous function equal to one on the boundary L and jumping to zero across its end points. Thus (2.118) follows from multiplication of end-points delta functions on δ_L according to rule (2.121).

A more illuminating example, given by the three dimensional singularity surface S with a density term which undergoes jump across some line $\ell \subset S$.



(Fig. 2.12)

A proof of (2.121) is greatly simplified by the introduction of the principal coordinates (u, v) on Fig. 2.12.

We now examine expression

$$\left\langle \overline{\frac{\partial g}{\partial u}} \cdot \delta_S, \varphi \right\rangle = \int_S \overline{\frac{\partial g}{\partial u}} \cdot \varphi \cdot dS$$

which can be rewritten as a double integral in curvalinear coordinates (u, v) with area element $dS = |G|^{1/2}$, obtained from the first fundamental matrix G:

$$= \iint_S \overline{\frac{\partial g}{\partial u}} \cdot \varphi \cdot |G|^{1/2} \cdot du \cdot dv$$

then according to FTED (2.119)

$$= \iint_S \frac{\partial g}{\partial u} \cdot \varphi \cdot |G|^{1/2}\, du\, dv + \int_\ell m_u \cdot \varphi \cdot [g] \cdot |G|^{1/2} \cdot d\ell$$

$$\text{Thus } \overline{\frac{\partial g}{\partial u}} \cdot \delta_S = \frac{\partial g}{\partial u} \cdot \delta_S + [g] \cdot m_u \cdot \delta_\ell \tag{2.122}$$

If we assume that g is constant inside and outside Ω on Fig. 2.12 and undergoes unit jump across its boundary ℓ, we have from (2.122) verification of the general rule (2.121).

$$\sigma \cdot \delta_L \cdot \delta_S = \sigma \cdot \delta_L$$

where density $\sigma = m_u \cdot [g]$ and m_u is the projection of the *tangential normal* m onto the u-direction.

Clearly, the above calculations do not represent the proof of more general identity (2.121). The rigorous theory should be extended even further to be able to multiply arbitrary order multilayers, for taking higher order generalized tangential derivatives, in the manner similar to derivatives of §2.9 in combination with tangential FTED (2.119), will run into this multiplication problem. However, we see already that conceptually, differentiation of the generalized functions with singular boundaries is no different than differentiation of distributions concentrated on smooth surfaces in the earlier considerations.

In view of the fundamental identity (2.121), the direct methods of the numerical interaction with singularities of the boundaries may be developed. Examination of the formulas for higher order derivatives in §2.9 shows that the same mathematical procedure which looks for breaks of the density term will be applicable to both finding breaks in the smoothness of the boundary as well as computing breaks in the smoothness of the jump function (and jumps of arbitrary order normal derviatives) along the boundary.

For example, if one is interested in finding points where curvature of the edge changes discontinuously, while the direction of the tangent is continuous function, generalized derivatives of at least *third order* should be combined accordingly to the kind of response desired. The reason can be seen from examination of formula (2.115). The term which carries information of the rule of change of the curvature there is $\delta_m \delta_i n_m$.

In particular, in case $\dfrac{\overline{\partial^3 f}}{\partial x^2 \partial y}$ is calculated and the boundary is planar, the contribution of the above term is

$$\delta_m \delta_i n_m = [f] \cdot \cos(\theta) \, (k \cdot \sin(\theta) - \dot{k}(\cos(\theta) + \sin(\theta)))$$

Hence in view of (2.121) and (2.115) if the curvature k undergoes jump $[k]$ on $P \in L$, i.e.

$$\dot{k} = \delta_{\ell} k + [k] \cdot \delta_P,$$

where δ_{ℓ} is regular tangential derivative in the direction of the tangent of L at P, (do not confuse it with the line delta function δ_{ℓ}) then $\frac{\overline{\partial^3 f}}{\partial x^2 \partial y}$ contains term $-[k] \cdot \cos(\theta)\,(\cos(\theta) + \sin\theta) \cdot \delta_P \cdot \delta_L$ which is a point distribution of the form $[k] \cdot \sigma \cdot \delta_P$. Similarly, contribution in $\frac{\overline{\partial^3 f}}{\partial x \partial y^2}$ may be computed and $\frac{\overline{\partial^3 f}}{\partial x \partial y^2} + \frac{\overline{\partial^3 f}}{\partial x^2 \partial y}$ is shown to have a term equal to $-[f] \cdot [k] \cdot \delta_P$. We can now detect such a point P by first convolving $f(\overline{x})$, with as small supported as possible, circular test function and then computing the above combination of generalized derivatives. Clearly, there are other choices for the *curvature jump detector* which would have to be tested experimentally. In particular, generalized gradient of the Beaudet operator, analyzed in (2.103) may happen to be even more effective. Its form is given as follows

$$|\overline{\nabla}\, B(f)| = \sqrt{\left(\frac{\overline{\partial} B(f)}{\partial x}\right)^2 + \left(\frac{\overline{\partial} B(f)}{\partial y}\right)^2} \tag{2.122.1}$$

The above methods are just as valid and applicable in analysis and detection of any singular behavior of k-dimensional submanifolds of the n dimensional singularity surfaces ($k < n$).

Thus to detect moving corners, which introduce tangential breakdown on the surface formed by the *moving edge* in (x, y, t) coordinate system, narrow *cylindrically* shaped test functions would have to be used in conjunction with second order time-space differentiation. In general, if we want to study higher dimensional singular surfaces, for instance in applications to motion detection, *all* of the above theory is applicable since it is developed for singular surfaces of the *arbitrary dimension.*

Finally a comment about why Beaudet operator [Bea.] analyzed in (2.103) is useful in extracting corners in images. We have concluded in §2.9 that B(f) is just a curvature sensitive functional, though it is known primarily as a "corner detector." But, in fact, it has already been experimentally noticed by [Bea.] that "SAD operator has been able to differentiate curved edges from straight edges," where SAD stands for nongeneralized use of B(f), i.e., simply computing $f_{xy}^2 - f_{xx} \cdot f_{yy}$.

If formula for the second order generalized derivative (2.95) is modified according to (2.119) and substitution is made in (2.95) replacing all tangential operators δ_j by their generalized version $\overline{\delta_j}$, then in case boundary is not continuous a term $\delta_k n_i \cdot [f] \cdot \delta_L$ in (2.95) will contribute component in the form

$$\sigma \cdot [n_i] \cdot \delta_P \cdot \delta_L = \sigma \cdot [n_i] \cdot \delta_P \tag{2.122.2}$$

where $[n_i]$ is a jump of the component of the normal vector at P and σ is proportional to the jump [f]. Hence it is not surprising at all that B(f) reacts strongly to the jumps of the normal of the singularity boundary, referred commonly to as "corners."

In the conclusion we remark that this paragraph is not meant to be an exhaustive study of the numerical analysis of the boundaries with singularities. On the contrary, it is just a starting point, an indication how truly powerful is apparatus of generalized functions in capturing essence of the whole spectrum of questions arising in the Numerical Analysis of Singularities.

§2.11. ***Tensor products of singularity detectors.***

As a result of the foregoing theoretical development, a genuinely novel concept of the *products of singularity* detectors is briefly introduced here.

Before the rigorous description is presented, a somewhat more intuitive discussion is useful.

The study of the preceding sections had inevitably culminated time after time with a singularity contribution estimate in the generic form $\int_L \left[\frac{\partial^K f}{\partial \mathbf{n}^K}\right] \cdot \sigma \cdot \frac{\partial \varphi^K}{\partial x^K} \cdot d\ell$, with density component σ possibly dependent on the jumps of the various higher order features, such as tangential and curvature discontinuities considered in §2.10 etc. Particularly simple form is taken by the singular contribution in the FTED (2.23), it is proportional to the $\int_L [f] \cdot \varphi \cdot d\ell$. Henceforth we shall call contribution in this form as *linear contribution* for it depends linearly on the jump of the examined function $f(\overline{x})$.

Whenever detection schema is built on the basis of the linear singular contribution, which includes *all* of the schemas considered so far, very unpleasant "side effects" come into existence. Thus fixing the length of the detecting "thin mask," will allow to have parasitic nonzero singular contribution when the mask is centered on the extension of the actual edge, i.e., outside of the singularity. Similarly if the jump function [f] changes sign across some point, as in the case of the vertices of the checker board pattern, the linear singular contribution might turn zero even though an edge is obviously present. We may also want to combine the outputs of singularity detectors of different nature, for example a corner detector and motion detector in the manner that their interaction does not introduce erroneous effects. Another serious problem which plagues all known detection schemas is breaking up of the continuous singular boundaries, or so called *streaking problem*. It therefore can be seen that on the one hand linear singularity detectors are not local enough, as in the vicinity of the edge end, but on the other hand these schemas are too local and are blind to anything that happens outside of the support of the detecting mask, hence streaking, checker board effect, etc.

Even more important is to observe that if the support of the detection mask matches *exactly* the local shape of the singularity, the linear contribution will be maximal. In this case, of course, the detection will give more precise measurements. However, it is generally true that singularity boundaries are not piecewise linear submanifolds, for they have nonzero curvatures. Hence it seems as if we cannot possibly guess a local shape of the singularity without actually detecting it. But, in fact, in what follows we will attempt to do just that.

To start answering the above questions, the arsenal of the concepts borrowed here from the distributions theory needs to be extended. In the following definitions *convolution* and *tensor product* of generalized functions are presented.

Definition 2.11.1.

Let $f(\overline{x})$ and $g(\overline{y})$ be generalized functions so that $f(\overline{x}) \in \mathcal{K}'_n$ and $g(\overline{x})$

$\in \mathcal{K}'_m$, then the unique distribution $f(\overline{x}) \otimes g(\overline{y}) \in \mathcal{K}'_{n+m}$ is defined by the rule

$$<f(\overline{x}) \otimes g(\overline{y}), \varphi(\overline{x}, \overline{y})> = <f(\overline{x}), <g(\overline{y}), \varphi(\overline{x}, \overline{y})> \qquad (2.123)$$

Clearly, tensor product of two regular functions is simply their regular product. In addition, (2.123) implies that if $\varphi(\overline{x}, \overline{y}) = \varphi_1(\overline{x}) \cdot \varphi_2(\overline{y})$, then

$$<f \otimes g, \varphi> = <f, \varphi_1> \cdot <g, \varphi_2> \qquad (2.124)$$

Furthermore, it is shown in the theory of generalized functions, that tensor product is a *commutative* and associative operation. And the following identity holds

$$D_x(f(\overline{x}) \otimes g(\overline{y})) = D_x(f(\overline{x})) \otimes g(\overline{y}) \qquad (2.125)$$

where D_x is any linear differential operator. Similarly multiplication by the test function $\varphi(\overline{x}) \in \mathcal{K}_n$ can be carried inside the tensor product:

$$\varphi(\overline{x}) \cdot (f(\overline{x}) \otimes f(\overline{y})) = \varphi(\overline{x}) \cdot f(\overline{x})) \otimes g(\overline{y}) \qquad (2.126)$$

Tensor product construction is utilized in the extending usual operation of the convolution to distributions.

Definition 2.11.2.

Let $f, g \in \mathcal{K}'_n$ and $\varphi \in \mathcal{K}_n$ then convolution product $f * g$ is given by

$$<f * g, \varphi> = <f \otimes g, \varphi(\overline{x} + \overline{y})> \tag{2.127}$$

It is also shown that generalized convolution satisfies the very same properties of the usual convolution $f*g(\overline{x}) = \int_{R^m} f(\overline{x} - \overline{t}) \cdot g(\overline{t}) \cdot d\overline{t}$ such as commutativity, associativity and differentiation property. As a consequence of the definition formula (2.127), convolution of the generalized function $f(\overline{x})$ with smooth test function $\varphi(\overline{x})$ takes particularly simple form:

$$f*\varphi(\overline{x}) = <f(\overline{t}), \varphi(\overline{x} - \overline{t})> \tag{2.128}$$

and is called *regularization* of $f(\overline{x})$ by $\varphi(\overline{x})$ since $f * \varphi \in C^{\infty}$, i.e., smooth function.

To obtain *quadratic nonlinear* edge detection schema, consider expressions in the form

$$Q(f)(\overline{x}) = <D^1_{\overline{y}} f \otimes D^2_{\overline{t}} f, \varphi_1(x, y) \cdot \varphi_2(y, t)> \tag{2.129}$$

where $D^K_{\overline{\alpha}}$ is some linear differential operator with respect to the variable $\overline{\alpha}$.

Particularly if φ_1 and φ_2 have form $\varphi_i(\xi, \eta) = \varphi_i(\xi - \eta)$, (2.129) may be transformed according to the regularization formula (2.128)

$$= <D_y f \otimes D_t f, \varphi_1(x - y) \cdot \varphi_2(y - t)>$$

$$= <\varphi_1(x - y) \cdot D_y f,\ D_t f * \varphi_2(y)>$$

or by the multiplication property (2.126) we have

$$= (D_y f \otimes (D_t f * \varphi_2)) * \varphi_1(\overline{x}) \tag{2.130}$$

Now, let us assume that D_α above is just partial differentiation operator $\frac{\partial}{\partial x_i}$ and that $f(\overline{x})$ is a regular function satisfying conditions of the FTED. We further set test functions φ_i in (2.129) to be narrow supported, so that regular contribution of any generalized function convolved with φ_i is negligible compared to the singular contribution (see (2.23)). Then with use of the FTED, (2.129) is reduced to

$$Q(f)(\overline{x}) = <\varphi_1(\overline{x} - \overline{y}) \cdot n_i \cdot [f] \cdot \delta_{L_y},\ (n_i \cdot [f] \cdot \delta_L) * \varphi_2(\overline{y})> \tag{2.131}$$

where L_y signifies that singularity lies in the $\overline{y}$-coordinate plane.

Upon using regularization property (2.128) and definition of the simple layer from §2.2 we obtain

$$= \int_L \varphi_1(\overline{x} - \overline{y}) \cdot n_i \cdot [f] \cdot \left(\int_L \varphi_2(y - t) \cdot n_i \cdot [f] \cdot d\ell_t\right] \cdot d\ell_y \tag{2.132}$$

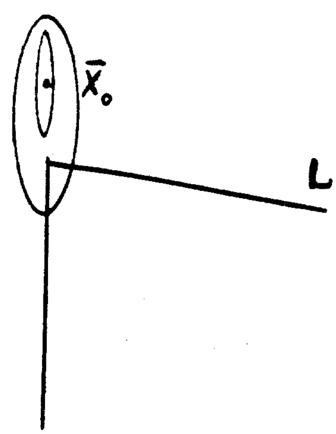
If we replace $\varphi_i(\xi)$ by its symmetric reflection $\psi_i(\xi)$ with respect to the origin, i.e. $\psi_i(\xi) = \varphi(-\xi)$, we get

$$Q(f) = <D_i f \cdot \psi_1,\ D_i f \odot \psi_2> \tag{2.133}$$

where ψ_1 and ψ_2 are sliding integration masks encountered in the earlier considerations and symbol $\odot$ denotes sliding integration, to distinguish it from the convolution.

From this last equation we can interpret result in (2.132) as a scheme whose singular contribution at the point is equal to the usual linear contribution with the density modified by the singular linear contribution of each point within the support of $\psi_1(\bar{y} - \bar{x})$. To rephrase the above statement differently, detection schema built on (2.133) detects submanifolds along which *linear detector* has significant output *on the average*.

It is worthwhile to notice that (2.133) is not commutative with respect to ψ_i's. Thus if ψ_1 and ψ_2 are two narrow test functions rotated by the equal amount and assuming supp(ψ_1) be shorter then supp(ψ_2) then for ψ_1 centered at $\bar{x}_0$ in the vicinity of the "corner" singularity L on Fig. 2.13.



(Fig. 2.13)

Q(f) will not give faulty measurements as would be the case if only "long" mask ψ_2 were used. The reason for this "smart" behavior is seen in (2.132), where ψ_2 measurement is multiplied by the local jump [f] within supp(ψ_2) which is *zero* outside L. On the contrary, if extension of the singularity is sought as in the striking problem, longer ψ_1 is chosen so that contributions picked by ψ_2 from far away points are integrated according to (2.132). Observe also that checker board problem, mentioned above, is also handled properly by the quadratic scheme (2.133). Since linear contributions do not vanish if mask is moved off the black-white vertex, our detector will simply disregard a single zero measurement and will integrate the rest according to (2.132).

The next task is to derive *computational form* of the simplest quadratic detector (2.133). By this is meant transformation of the *functional* form of Q(f) into the form when all terms resulting may be computed by the computer. To this end, we simply apply integration by parts several times to obtain expression.

$$Q(f) = \left[\psi_1 \odot ((f \odot \psi_2)_{x_i} \cdot f)\right]_{x_i} - \psi_1 \odot (f \cdot (f \odot \psi_2)_{x_i x_i}) \qquad (2.134)$$

where subscript is partial differentiation symbol and $\odot$ stands for the process of sliding integration. It is to be noted that Q(f) in (2.134) can be now numerically estimated with the use of numerical integration and differentiation. It would also be legitimate to take another, computationally more expensive form of Q(f), whose significance is in the new interpretation assigned to quadratic schemas. We may also write

$$Q(f) = -f \odot (\psi_1 \cdot (f \odot \psi_2)_{x_i})_{x_i} \qquad (2.135)$$

which can be thought as familiar linear detection with *modified* mask $\psi_1 \cdot (f \odot \psi_2)_{x_i}$ whose shape changes from point to point, according to the measurements of the *first pass* detection $f \odot \psi_2$.

Thus (2.135) represents space variant *nonlinear* filtering, correlating singular part of the actual generalized derivative of $f(\overline{x})$ with measurements of the detecting mechanism on the previous pass. This observation yields an *iterative* detection schema

$$Q_{n+1}(f) = -f \odot (\psi_1 \cdot Q_n(f))_{x_i} \tag{2.135.1}$$

Such a schema amounts to the molding "custom made" detection mask, which fits local shape and size of the singularity within its support. Normalization of each successive $Q_n(f)$ would be necessary to prevent numerical blowup.

A further improvement of this schema is made by observing that differential operators D^K in the definition (2.129) of the quadratic detector need not be just first generalized derivatives and need not be the same. Specifically, this generalized differential operator may be linear combination given by (2.100) of the first order detector and *curvature* detector developed in §2.9. Thus if x and y are directions of differentiation in (2.100) we may derive similarly to the (2.134), via integration by parts, computational form of the *second order* quadratic detector

$$Q^{(2)}(f) = (\psi_1 \odot (G \cdot f - G_y \cdot f))_x + (\psi_1 \odot (f \cdot (G_{xy} - G_x)) + \tag{2.136}$$
$$+ (\psi_1 \odot (G \cdot f))_{xy} - (\psi_2 \odot (G_x \cdot f))_y$$

where $G = (f \odot \psi_2)_x + \lambda \cdot (f \odot \psi_2)_{xy}$ is familiar edge-curvature detector.

Looking at quite cumbersome expression (2.136) one may find satisfaction in the fact that no intuition would produce it without the help from the calculus of generalized functions. Thus one more time underscoring its significance in the numerical analysis of singularities.

It is worthwhile to note that taking *third order and higher* tensor products, whose computational form may be derived just as easily as for quadratic detector (2.129), information from more and more distant singular neighborhoods will be integrated by the locally fixed detector. In such a way, *global* singularity detection can be set up. Another important observation is that the whole detection schema may be fashioned so that the mask correction term in (2.135) comes not from the directional measurement provided by the integration with the single directed mask ψ_1 on the previous iteration, but from the compound procedure which combines all directional outputs from the preceding iteration. Such a procedure is described briefly in the next paragraph, containing some preliminary experimental results.

§2.12. *A simple edge detector.*

Before proceeding with the business of the description of the experiments with "narrow rotating masks" algorithm, *formulated* in this work, the following disclaimer is called for by the emerging view of the nature of Numerical Analysis of Singularities:

*The author of this work does not believe that this new numerical discipline should focus primarily on the "**detection**" of singularities, i.e., just finding loci of the singularity boundaries. But numerical methods should be developed whose purpose is **computation** of the essential singularities.*

And what is there to compute besides the coordinates of the singularity boundaries? All these *internal* numerical quantities, which come into being only on the surfaces of singularities. That such quantities actually exist have been assured, I hope, by the preceding calculations based on the theory of generalized functions.

On the whole, it looks that to solve the *singularity location* (i.e. detection) problem which today's AI applications seem to demand, some inner characteristics of this singularity (curvatures, for example) will have to be computed as well. In this sense, the so called Fitting approach pioneered by M. Hueckel [Hu.] and continued recently in [Na. Bi.] is much closer to the spirit of this work. However, from the point of view of the richness of the computational structure of the essential singularity, only primitive first order approximation has ever been attempted there. It hardly needs to be said, therefore, that efforts in this field should be refocused from the "detecting" attitude which has been a dead end for 30 years, to the numerical analytic orientation.

The remainder of this chapter is devoted to the discussion of the experimental *edge* detector (notice by FTED edge is a particular kind of singularity), which combines first order detector and three different curvature detectors with the method of rotating thin masks described previously. These measurements are then combined into the compound edge map of variable intensity through the variant of nonmaximal suppression scheme, and then image is further nonmaximally suppressed in the estimated gradient direction.

The author has critical reservations about the very nature of this passage from the local measurements to the global singularity description. But until less ad hoc methods are developed, this is state of the art which is adapted as well as we could to the particulars of the rotating thin masks detector. While resulting edge detecting appears superior in many respects to the previously published experimental data (of course the author does not have information if any classified data of better quality exists), we have used the final edge extraction just to illustrate that even this faulty approach (i.e., nonmaximal suppression) can be improved significantly by incorporating more sensible local singularity measurements. Unfortunately experimentation has been limited for the lack of high speed interactive imaging facilities usually available in today's AI oriented laboratories.

The first step of our detection scheme is to choose appropriate *scale* according to the scale of the detail sought and to the estimated noise level. This is accomplished by computing smooth test functions, or mollifiers as they are called in functional analysis, in the form

$$\varphi_{MN}(x, y) = \begin{cases} \exp\left(-\frac{M^2}{M^2 - x^2}\right) \cdot \exp\left(\frac{N^2}{N^2 - y^2}\right) & \text{if } |x| \leq M \text{ and } |y| \leq N \\ 0, & \text{otherwise} \end{cases}$$

M, N are integers which determine "length" and "width" of $\varphi(x, y)$.

Clearly $\varphi(x, y)$ is a smooth function with compact support akin to the test function defined in (2.1.0). If we set the step of the digital grid to be a unit, then $\mathrm{supp}(\varphi)$ is $(2M + 1) \cdot (2N + 1)$ digital rectangle.

Next, the *angular resolution* K is chosen and the function $\varphi(x, y)$ is rotated K times through the angle $\alpha = \pi/K$. After each rotation, the digital version $\varphi_{M \cdot \alpha}(x, y)$ is recomputed by simply rotating every grid point (x, y) in $\mathrm{supp}(\varphi_{n \cdot \alpha})$ backward using transformation formulae

$$\begin{aligned} X' &= \cos(m \cdot \alpha) \cdot x + \sin(m \cdot \alpha) \cdot y \\ Y' &= -\sin(m \cdot \alpha) \cdot x + \cos(m \cdot \alpha) \cdot y \end{aligned} \tag{2.138}$$

and then recomputing corresponding function value from the original test function in (2.137), i.e.

$$\varphi_{m \cdot \alpha}(x, y) = \varphi_{M,N}(X', Y') \tag{2.139}$$

Thus by having *explicit* form of the test function we avoid interpolation problems associated with rotation of the digital masks.

Picture 2 shows resulting test functions, where the intensity of the pixel correspond to the value of the function at this point. The angular resolution on (Pict. 2) is 15°. Each row of the masks has a scale associated with it. Thus in the first row $M = 2$, $N = 4$, so the masks are 2 pixel wide and 8 pixels long, and in the last row we have 8×8 pixels masks (i.e., $M = N = 4$). Hence from row to row we have transition from strongly directed masks to less directed.

The next step of the algorithm is digital integration of the given image with the sequence of sliding masks of fixed scale and fixed angular resolution. Which means that for each oriented function $\varphi_{m \cdot \alpha}(x, y)$ and image $f(x, y)$ a new function $f_{m \cdot \alpha}$ is derived according to

$$f_{m \cdot \alpha}(x, y) = \iint_{B_\epsilon} f(\xi, \eta) \cdot \varphi_{n \cdot \alpha}(\xi - x, \eta - y) \cdot d\xi \cdot d\eta \tag{2.139}$$

where B_ϵ is support of the shifted $\varphi_{n \cdot \alpha}$ and double integral in (2.139) is numerically evaluated by your favorite quadrature method. This process makes K differently blurred versions of the original image, where π/K is angular resolution.

We now choose *central* finite differences to approxiamte digital derivatives:

$$\Delta_x f_{i,j} = \tfrac{1}{2}(f_{i+1,j} - f_{i-1,j}) \tag{2.140}$$

Derivative in the direction $\ell = m\alpha$ is computed according to formula

$$\Delta_\ell = \cos(m \cdot \alpha) \cdot \Delta_x + \sin(m \cdot \alpha) \cdot \Delta_y \tag{2.141}$$

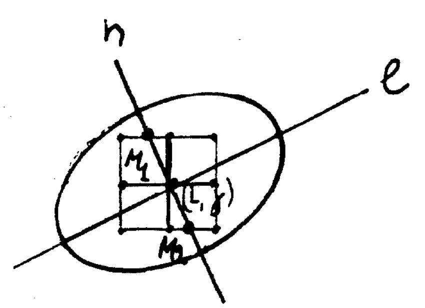
Approximation to the second derivatives in the direction ℓ is obtained by simply iterating operator Δ_ℓ twice, i.e.,

$$\frac{\partial^2 f}{\partial \ell^2} = \Delta_\ell(\Delta_\ell f) = \Delta_\ell^2 f \tag{2.142}$$

and similarly for mixed derivatives

$$\frac{\partial^2 f}{\partial \ell \partial n} = \Delta_\ell \Delta_n(f) \tag{2.143}$$

where n, ℓ are some directions of differentiation. Let n be the direction associated with positive direction of the short axis of the masks on (Pict. 2) and ℓ is direction of its long axis as on Fig. 2.14.



(Fig. 2.14)

Then for each $f_{m\cdot\alpha}$ we proceed to apply three following finite difference operators:

$$D_1(f_{m\cdot\alpha}) = \Delta_n f_{m\cdot\alpha} + \lambda \cdot \Delta_n\Delta_\ell f_{m\cdot\alpha} \tag{2.144}$$

$$D_2(f_{m\cdot\alpha}) = |\Delta_n f_{m\cdot\alpha}| + \lambda \cdot |\Delta_\ell^2 f_{m\cdot\alpha}| \tag{2.145}$$

$$D_3(f_{m\cdot\alpha}) = \Delta_n f_{m\cdot\alpha} + \lambda \cdot |(\Delta_n \Delta_\ell f_{m\cdot\alpha})^2 - (\Delta_\ell^2 f_{m\cdot\alpha}) \cdot (\Delta_\ell^2 f_{m\cdot\alpha})| \tag{2.146}$$

The reader, of course, will recognize these as a mixture of the first order detector with various forms of the curvature detectors described in §2.9. Constant λ determines the weight ascribed to the curvature of the singularity.

We demonstrate this process thoroughly for the 256×256 "clock" image on (Pict. 3), which was chosen for its great deal of small details, including easily perturbed numerals on the "clock's" face. It will give us some idea how well the schemes extract singular information. The other reason for such detailed picture exposition is our effort to avoid unfortunate practice in the literature on edge detection, when only a few pictures of very questionable *informational content* are given, usually the ones which authors consider "the best presentable" and most publishable.

In all images therein, integration was performed with the masks of the first row of (Pict. 2) whose size is 2×8 pixels and $\alpha = \pi/12$. Picture 4 (i) depicts on the left the result of the integration with mask #3 and on the right, corresponding $D_1(f_{m\cdot\alpha})$ of (2.144) is evaluated and mapped as an intensity function. Furthermore (Pict. 4) (ii) does the same for mask #7.

Similarly, (Pict. 5)–(Pict. 6) demonstrate this process for the digital differential operators D_2 and D_3 of (2.145), (2.146).

As a result we have K processed images $f^{\nu}_{\alpha \cdot m}$, $m = 0, 1, \ldots, K - 1$ for each differential operator D_{ν} in (2.144)–(2.146). This completes stage of the *local detection*.

For each (i, j) point and fixed ν, the $\max_{m} (f^{\nu}_{m \cdot \alpha}(i, j))$ is computed, say it is $f^{\nu}_{\alpha_0}(i, j)$ on Fig. 2.14. We interpolate values $f_{\alpha_0}(M_i)$ of the intersection points with the local grid lines in the mask's direction **n**. First order interpolation has been used here. Then condition

$$f_{\alpha_0}(i, j) \geq f_{\alpha_0}(M_i)_{,i=1,2} \tag{2.147}$$

is checked. If (2.147) is true then we put $F_{\nu}(i, j) = f^{\nu}_{\alpha_0}(i, j)$; $\nu = 1, 2, 3$. If inequality is not satisfied then sorting continues for the rest of $f^{\nu}_{\alpha m}$ and the maximal remaining element is checked for (2.147) again, etc., until we exhaust all possibilities and then set $F_{\nu}(i, j) = 0$.

Images F_{ν}, $i = 1, 2, 3$ are shown on (Pict. 7) (i), (ii), (iii), correspondingly. Curvature weight $\lambda = 1$ has been used there. We call this variable intensity edge map a "*pencil drawing*," for its obvious resemblance.

Some noted differences among F_{ν} can be observed. For example, F_1 has the best "clock" outline and most of the numerals are easily discernible, while F_2 and F_3 seem to get better such highly curved characters as numeral "6." Smaller values of λ have been tested with inevitable "erosion" of all curved edges. Clearly, thresholding is not a clever thing to apply to the informationally reach image F_{ν}, and some other global methods should be used to extract bilevel edge map (for example a variational approach).

Nonetheless, we perform direct thresholding of F_{ν}'s on the same level just for the sake of comparison. Again we see how well curvature is captured by D_1 operator. See (Pict. 8) for these thresholds.

We can further improve this ad hoc suppression scheme by "thinning" the lines in F_{ν} to 1 pixel wide. For each $f_{\nu}(i, j)$ the direction of the gradient vector is estimated by computing at each point (i, j)

$$\mathbf{grad}(F_{\nu}(i, j)) = (\Delta_x^+ F_{\nu}(i, j), \Delta_y^+ F_{\nu}(i, j) \qquad (2.148)$$

where Δ^+ is forward difference operator. Then the same interpolation procedure as above is performed and (i, y) is checked for maxima condition. If successful, the value is preserved. If not, the value is suppressed, i.e., turned zero. The results are shown on (Pict. 9), and corresponding equal thresholds on (Pict. 10).

It is worthwhile to see the same detection process for the quite noisy 512×512 image of the Tank (Pict. 11.1). On the whole, D_1 operator seems to have the best performance, which is depicted on the sequence Pict. 11. Again we see that every minute detail of the "tank" image shows up in the "pencil drawing" image, including tiny treads. We also see how indiscriminately thresholding destroys singularity information, so painstakingly obtained with the use of directional integration.

Finally we present 512×512 computer generated "Robot" image (Pict. 12.1)

The interesting observation about this image is the presence of *bright* and *dark* regions containing singular boundaries. According to the FTED and subsequent curvature calculations, the singular contribution, which we are observing as an intensity of the obtained “pencil drawing” image, is proportional to the size of the jump. Since this quantity is small in dark regions relative to the jump measurements of bright regions, we may lose these “dark” singularities.

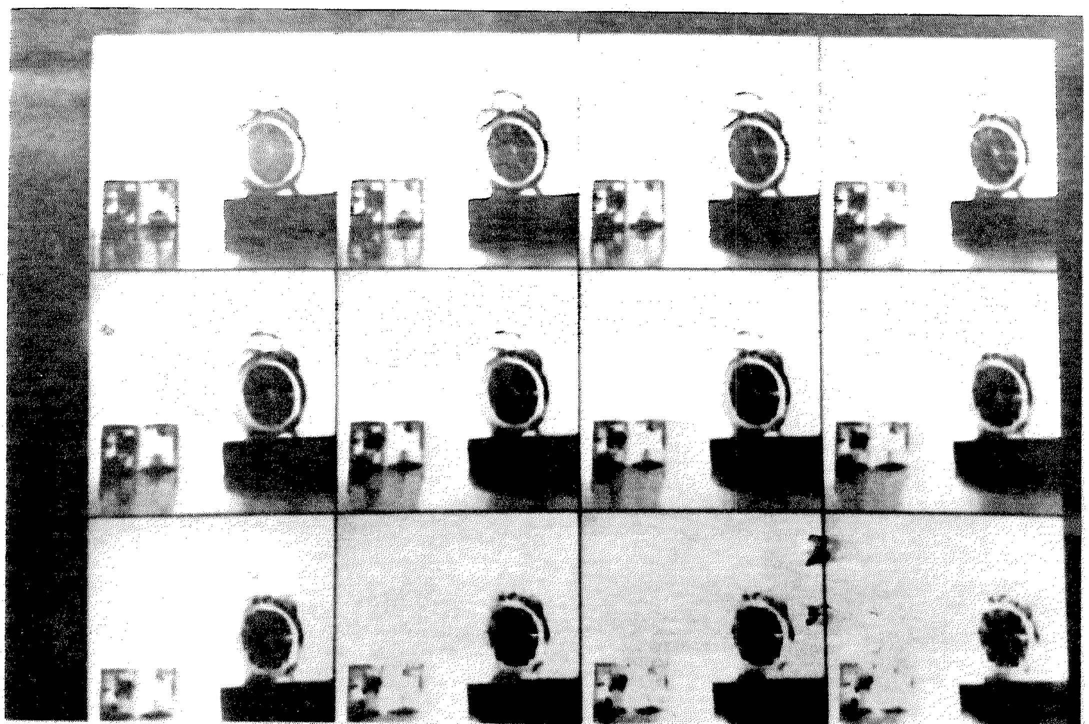
To overcome this problem we modify each $f^{\nu}_{m\cdot\alpha}(x, y)$ by dividing it by the local directed energy estimate $E = \|f_{m\alpha}(x, y)\|_2$, where for each point $(i, j) = (x_0, y_0)$

$$E(x_0, y_0) = \sqrt{\iint_{B_\epsilon} (\varphi_{m\cdot\alpha} \cdot f)^2 \cdot dx \cdot dy} \tag{2.149}$$

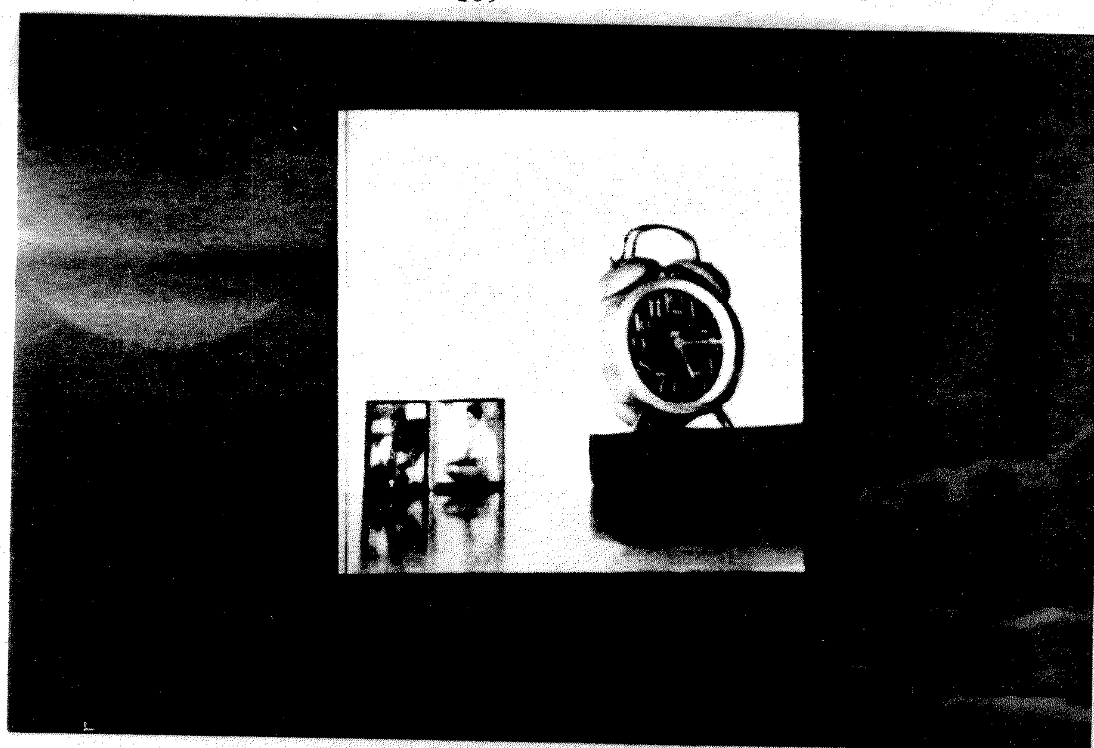
where B_ϵ is support of the $\varphi_{m\cdot\alpha}$ centered on (x_0, y_0).

Using this local energy equalization with the developed above scheme, we obtain sequence of images depicted on Pict. 12. Observe how well it performs in detecting almost invisible features from the original image (Pict. 12.1).

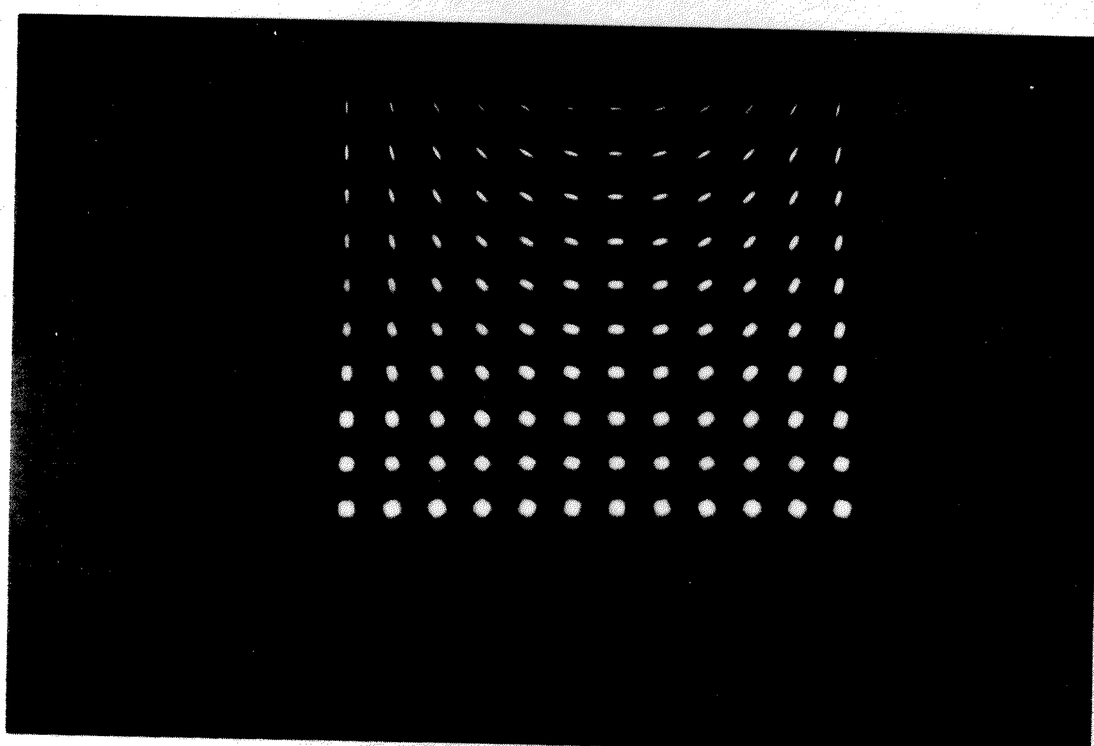
These numerical experiments shall continue with a variety of other more sophisticated local detection shemas, described in this work. Of course, it will include 3-dimensional singularity surfaces with torsion, produced by *motion sequences with possible rotations*. On the theoretical side, construction of these schemas is not that different, since the same theory developed here is applicable. However, this will require much more powerful interactive computing facility than available to the author at the moment.



"Clock 1"

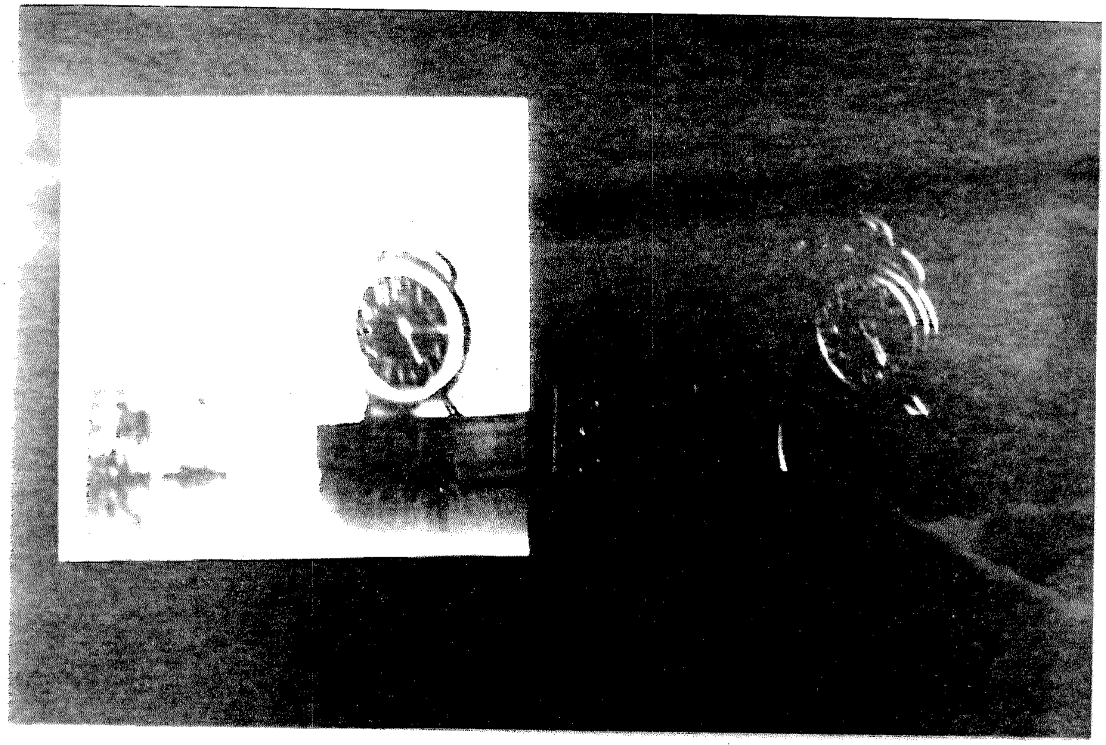


(Pict. 3)

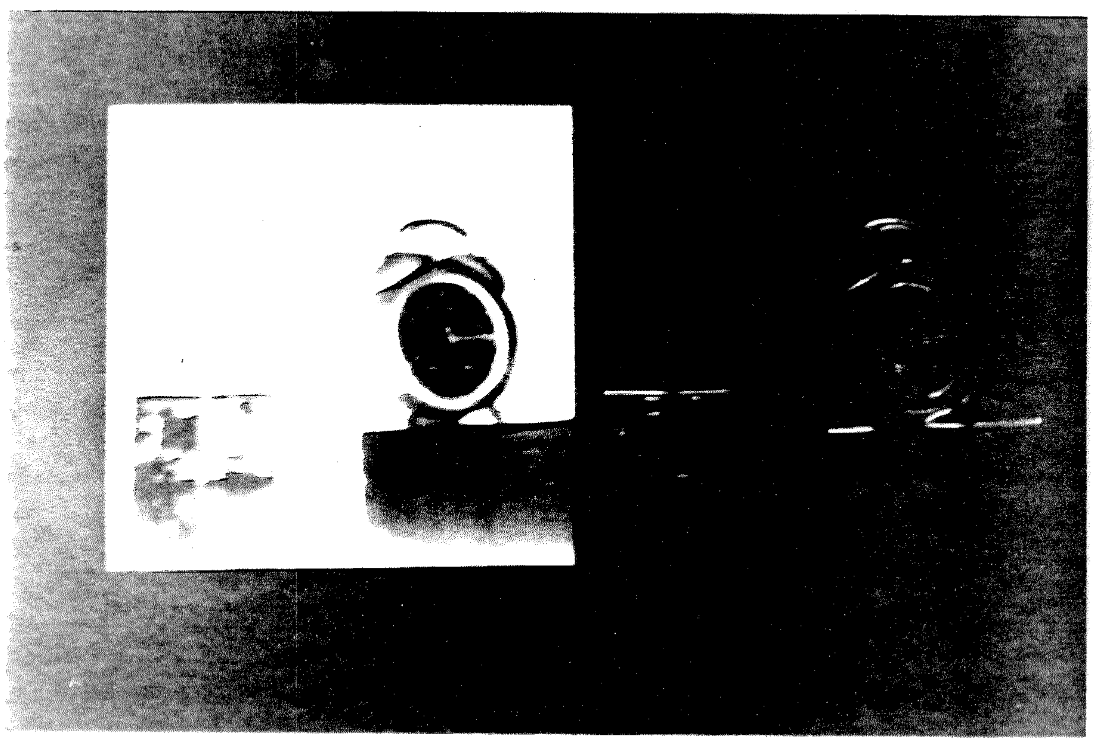


(Pict. 2)

(Pict. 4)

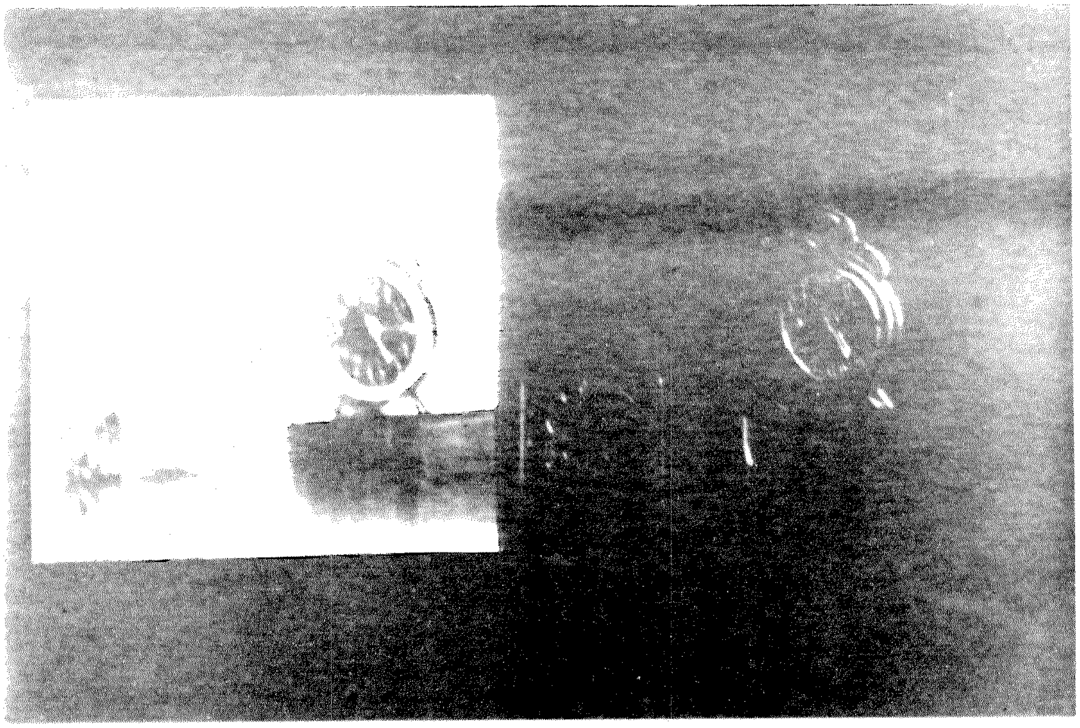


(i)

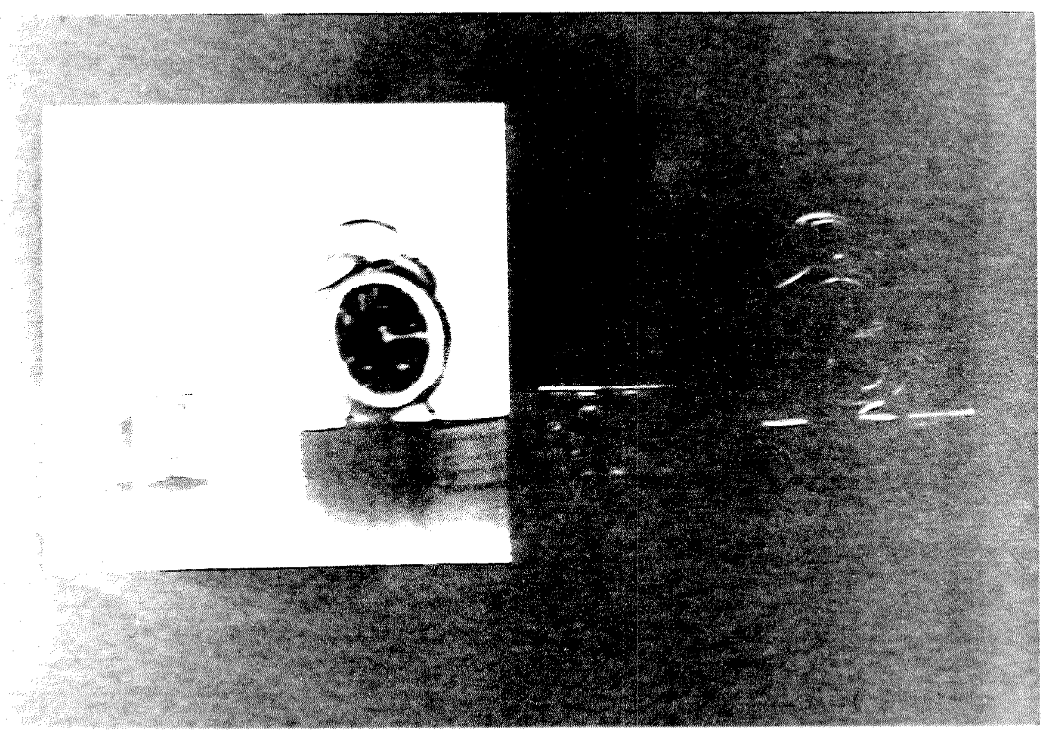


(ii)

(Pict. 5)



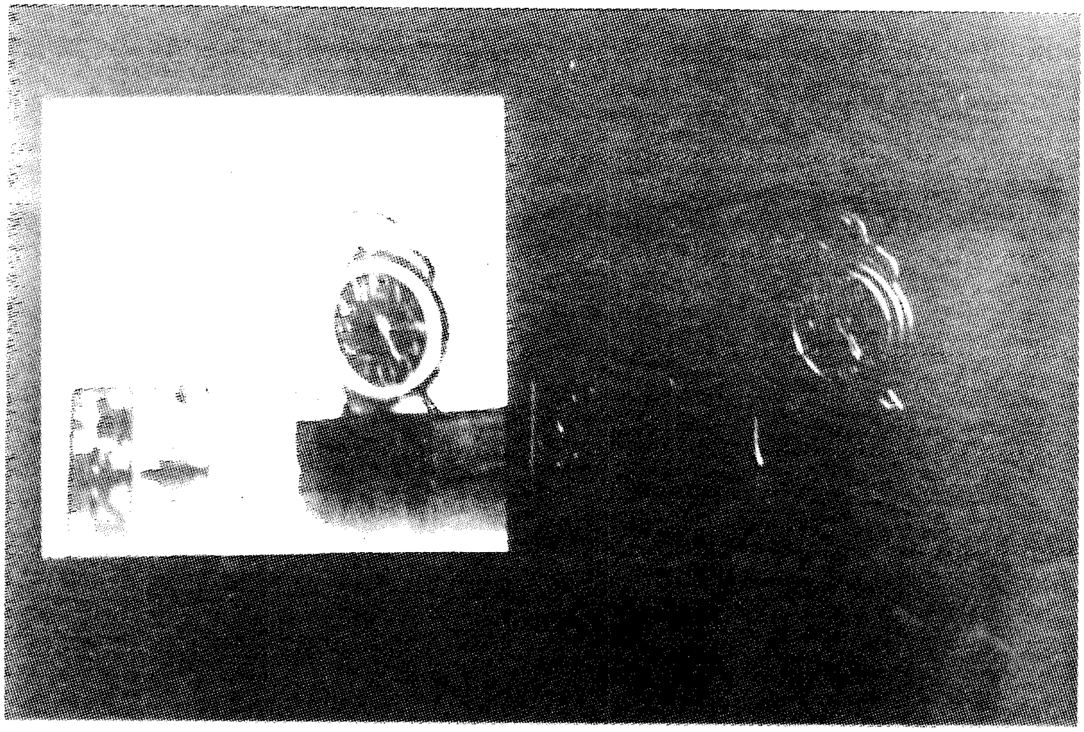
(i)



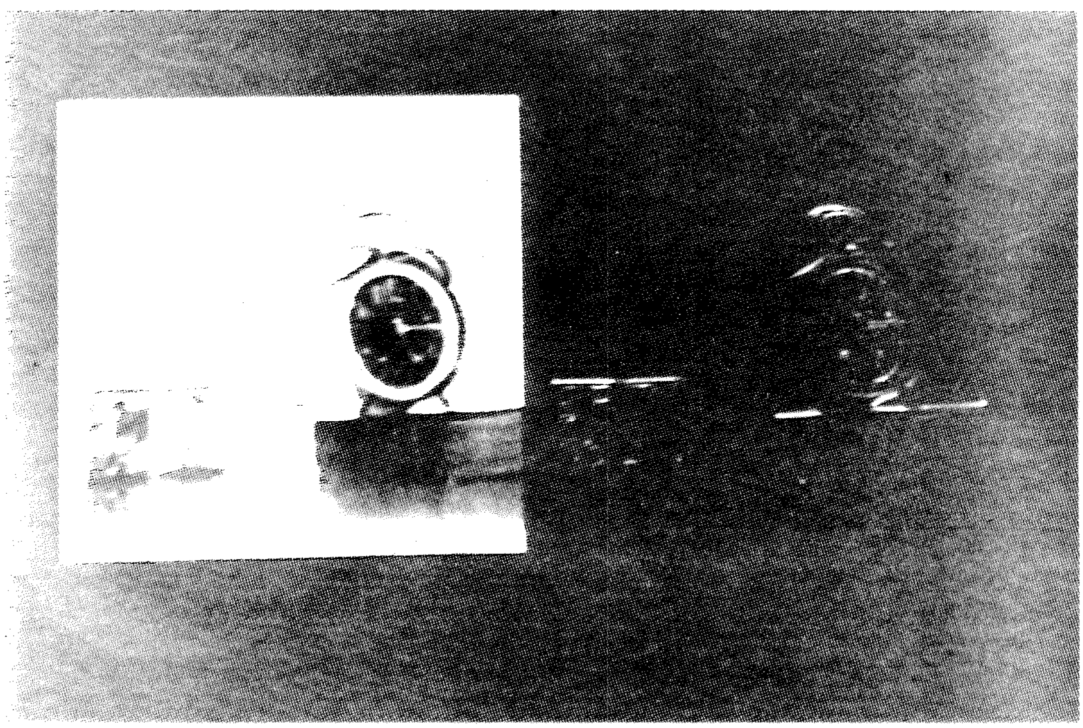
(ii)

(Pict. 6)

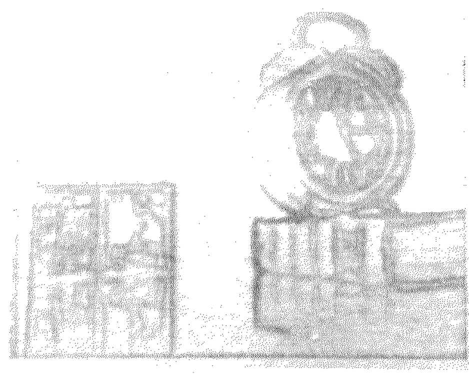
(i)



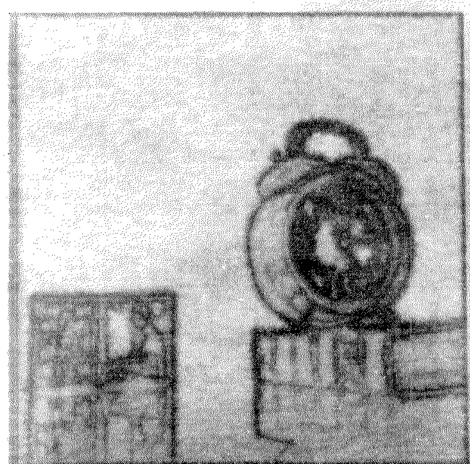
(ii)



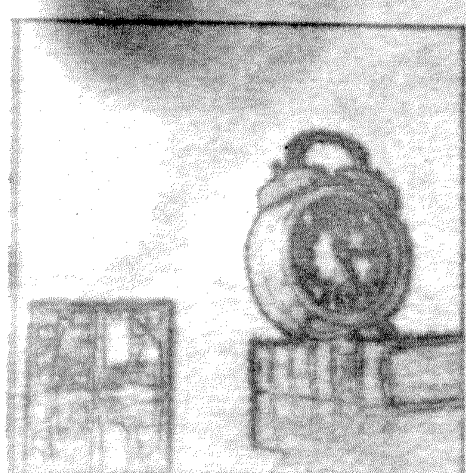
(Pict. 7)



(i)

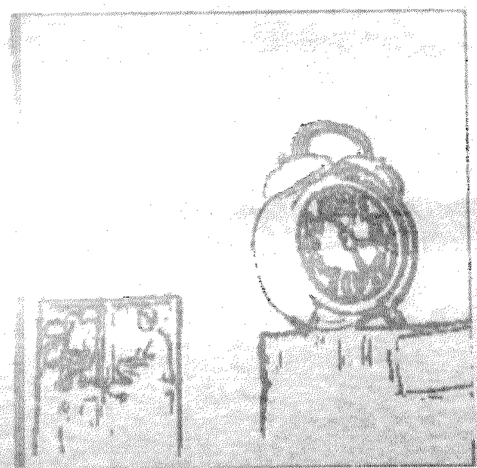


(ii)

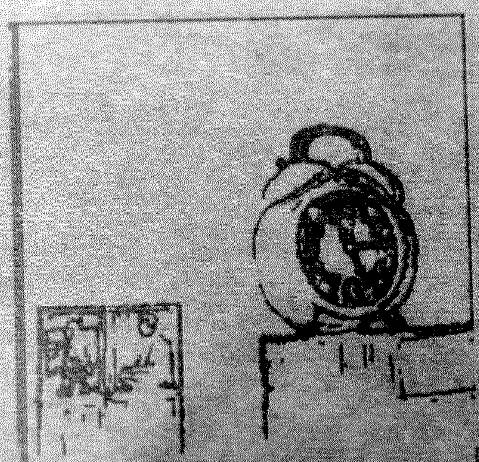


(iii)

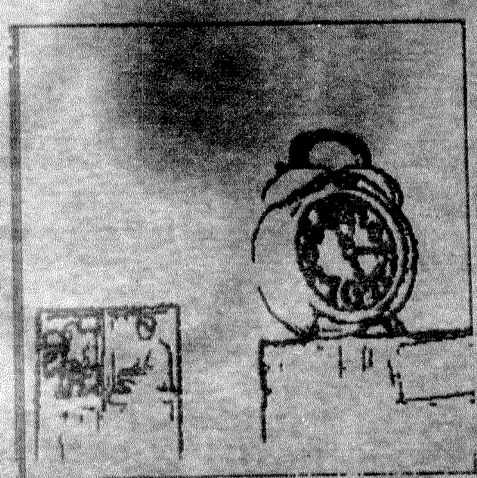
(Pict. 8)



(i)

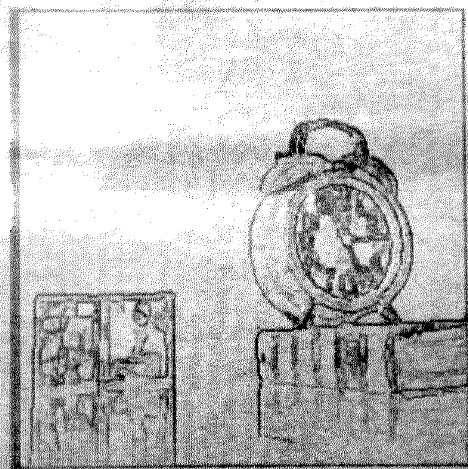


(ii)

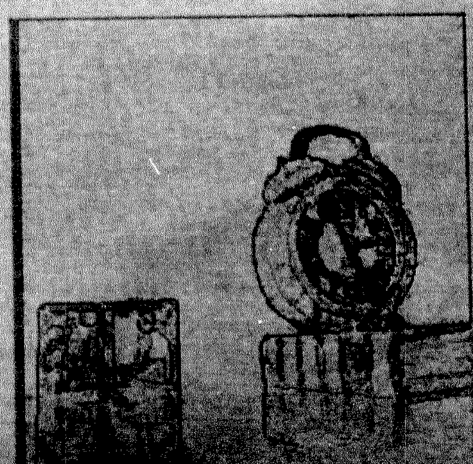


(iii)

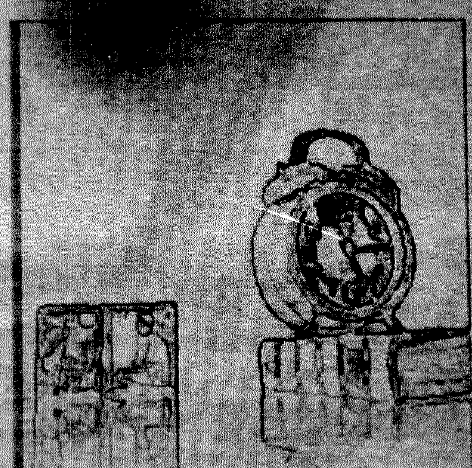
(Pict. 9)



(i)

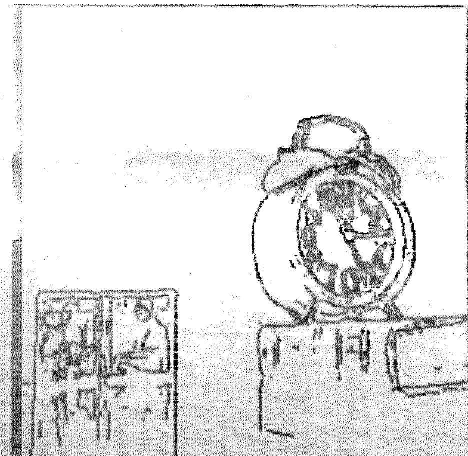


(ii)

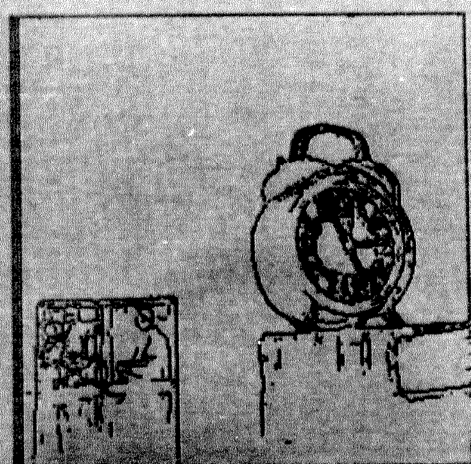


(iii)

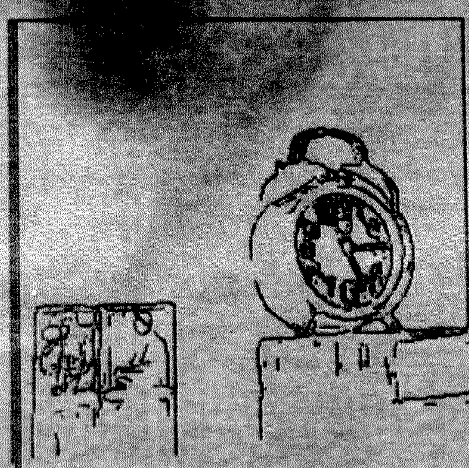
(Pict. 10)



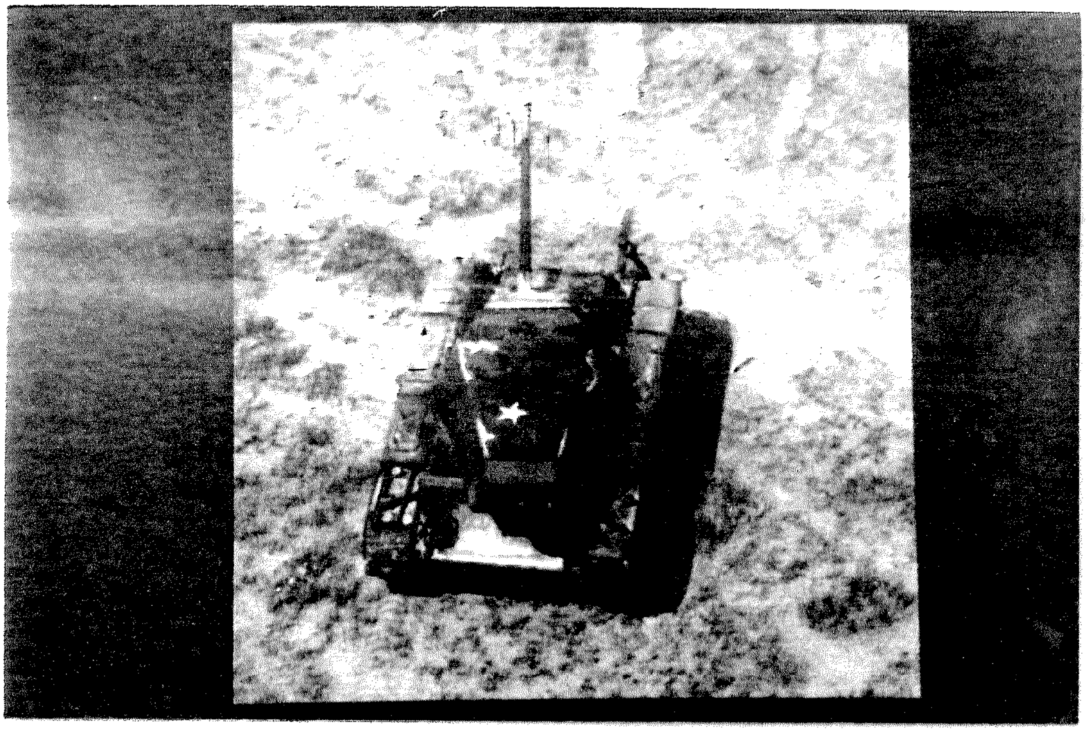
(i)



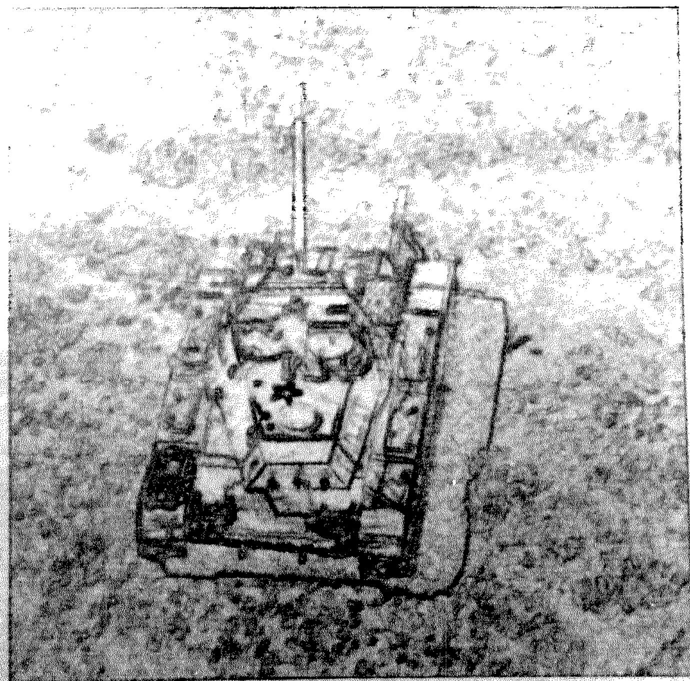
(ii)



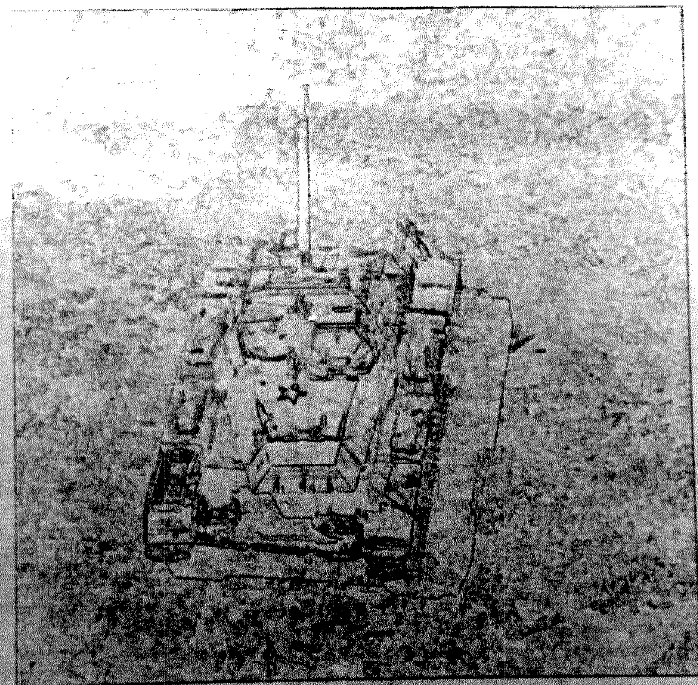
(iii)



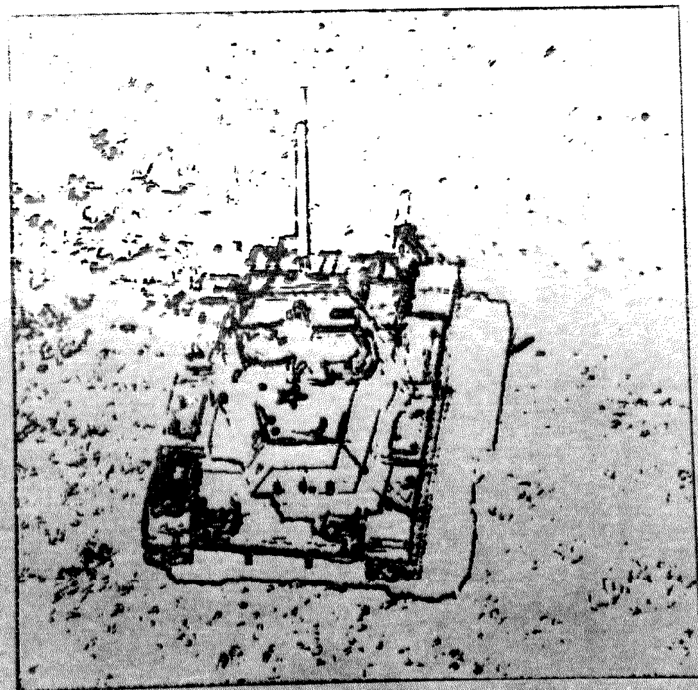
(Pict. 11(i))



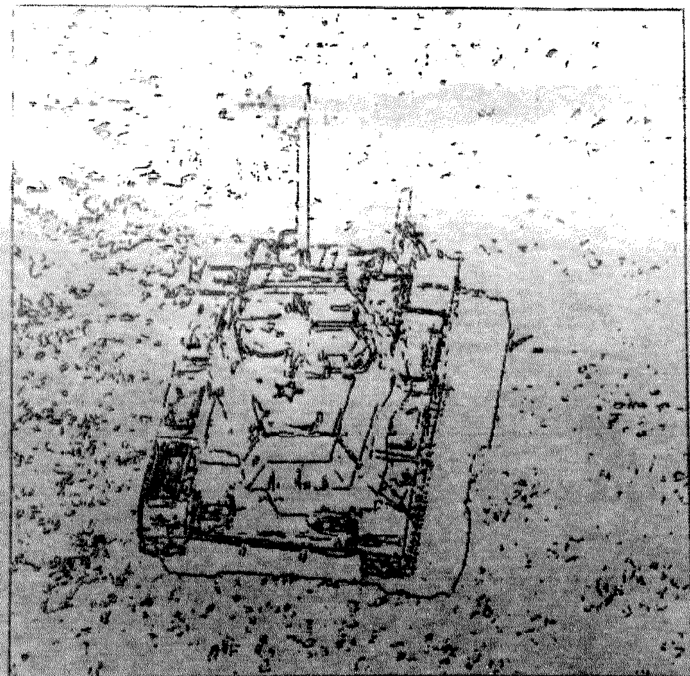
(Pict. 11(ii))



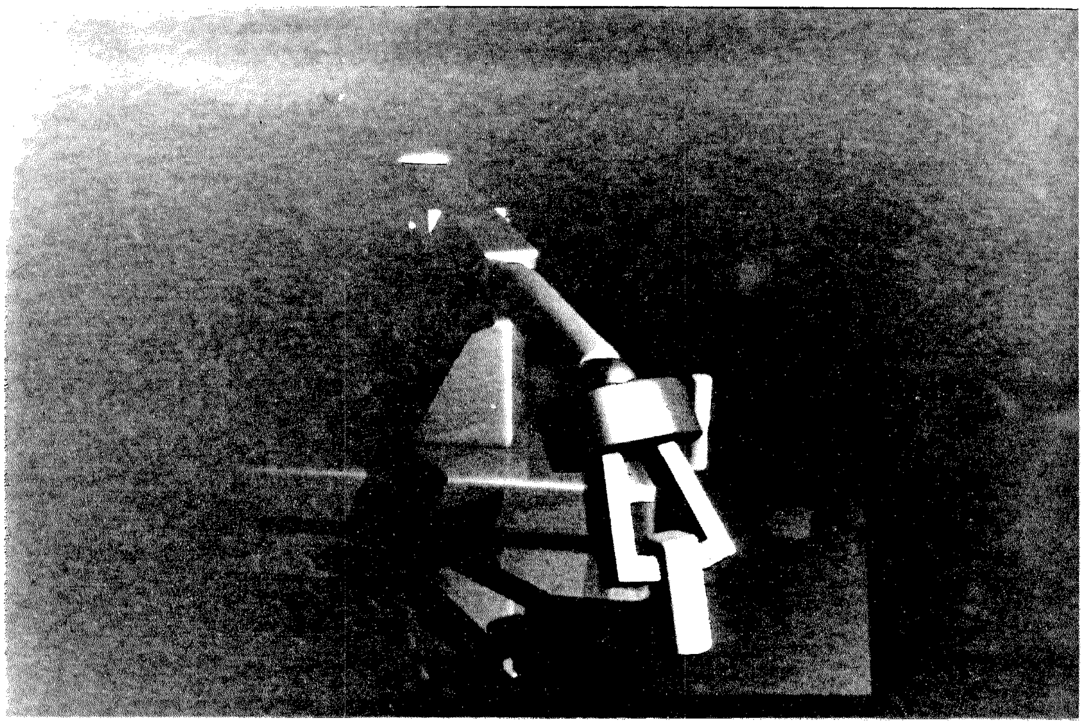
(Pict. 11(iii))



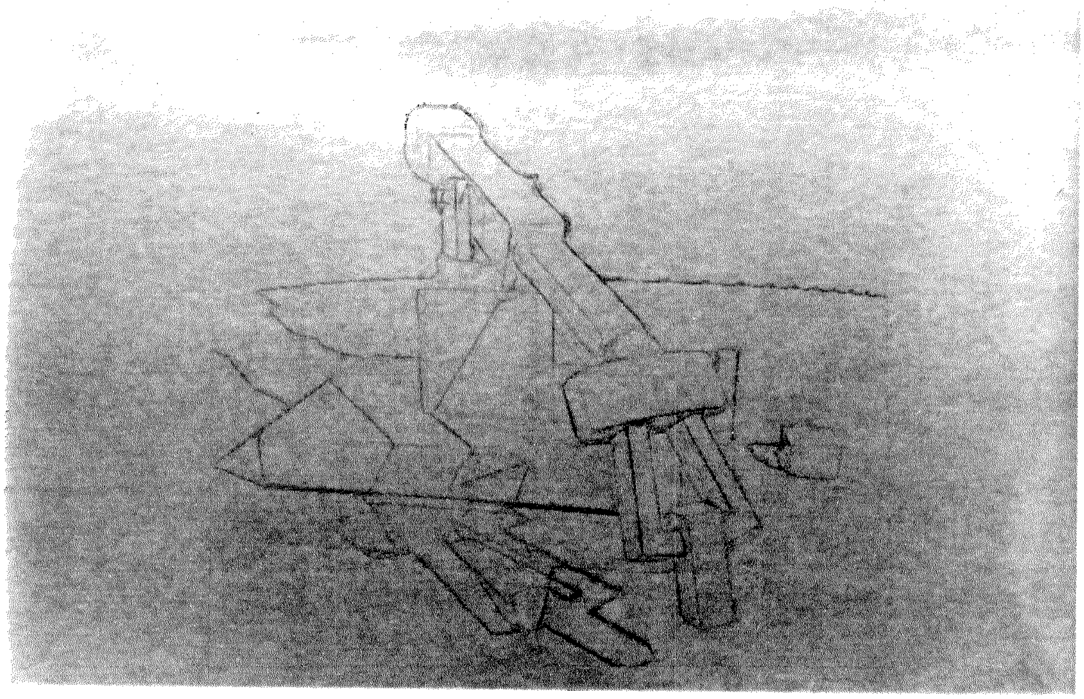
(Pict. 11(iv))



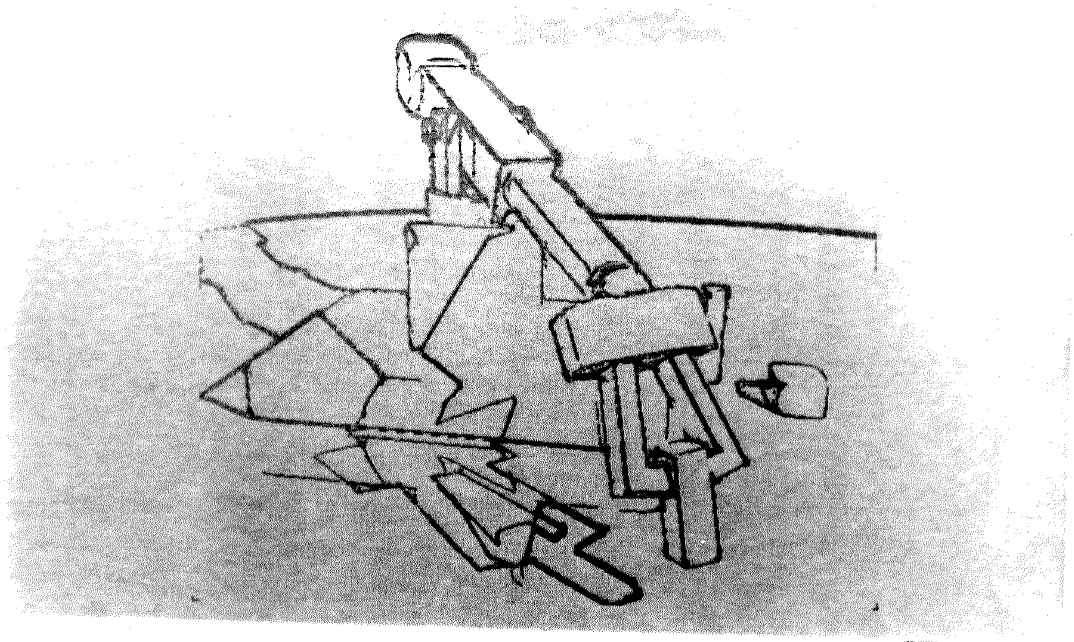
(Pict. 11(v))



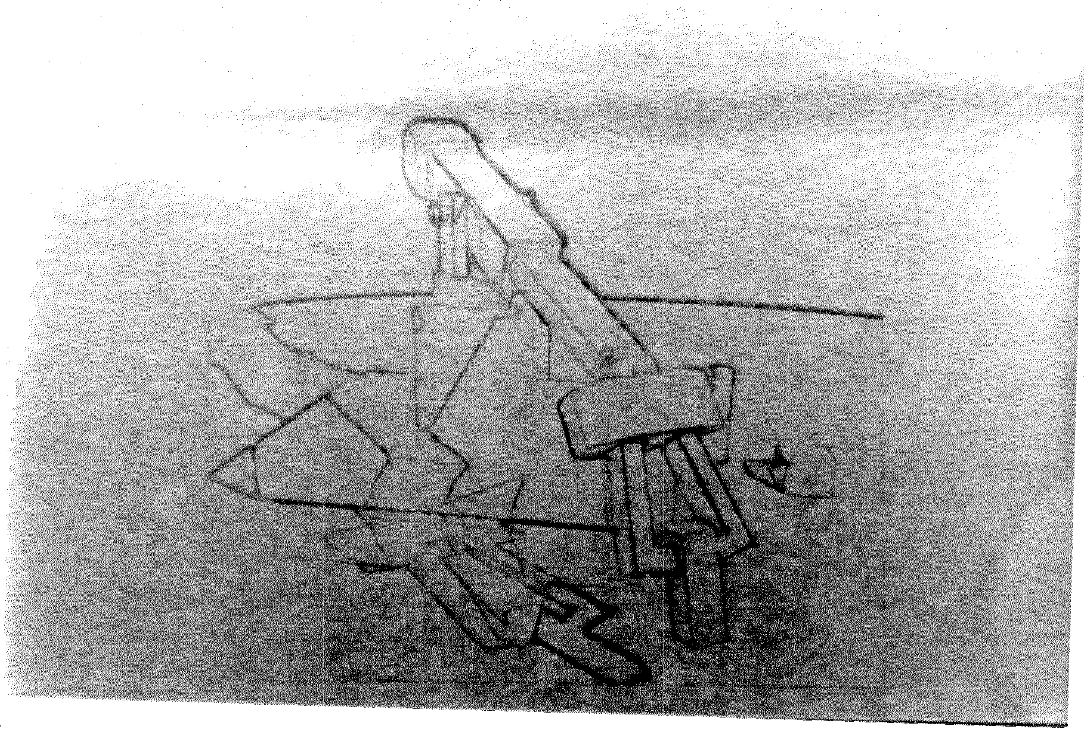
(Pict. 12.1)



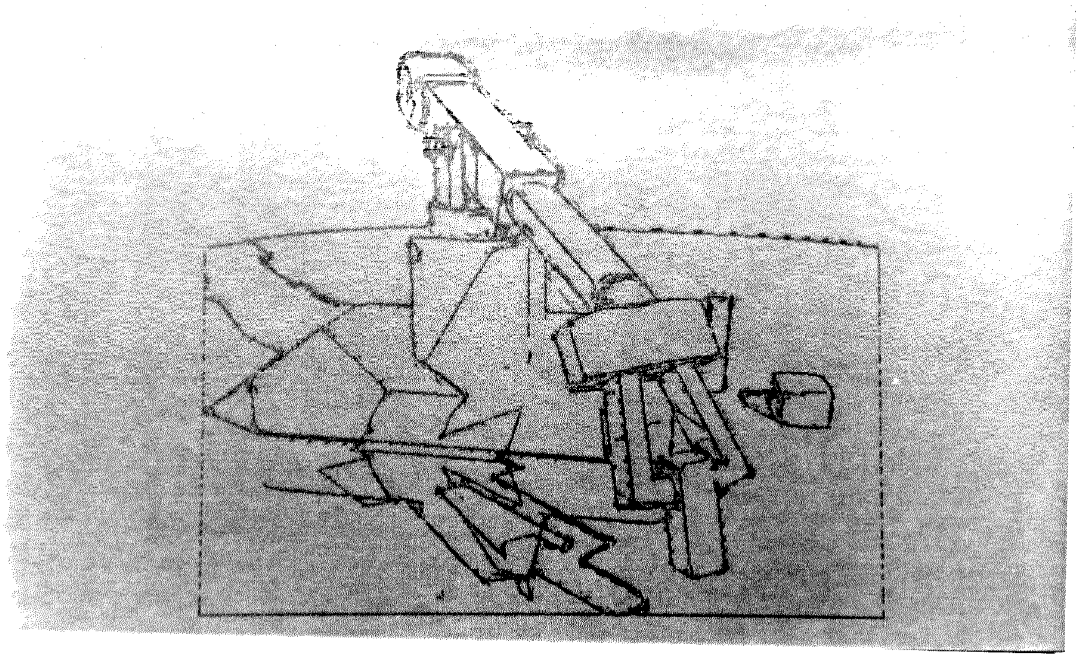
(Pict. 12.2)



(Pict. 12.3)



(Pict. 12.4)



(Pict. 12.5)

§2.13. *Remarks on the consequences of discrete sampling.*

The distributional approach developed in this work represents only the beginning fragment of the Numerical Analysis of Singularities.

With few exceptions, the generalized functions, whose essential singularities we examined, were assumed to be given by continuum description. However in the formulation of the "narrow rotating masks" algorithm of §2.12 we have switched to the discrete numerical implementation manifested in the discrete sampling of the image, discrete integration procedure, finite-difference approximations and finally non-maximum suppression based on discrete interpolation of the computed generalized derivatives.

Each of these steps introduces an approximation error. Theoretical estimation of the committed error and its influence on the resulting singularity detection is of great interest. Another source of errors is a noisy input data, whose effect is known to destroy outputs of many edge detecting schemes. For example, well performing in the absence of noise Sobel, Compass and Roberts edge detectors are unstable even with high signal to noise ratios, e.g., SNR — 10 [Pra].

The effect of noise on the localizing ability of a one-dimensional linear edge detector was considered in [Can.]. First, it has been shown there that there exists an optimally shaped edge detector with respect to the signal to noise performance as well as localization property. This optimal one-dimensional shape is derived through the numerical optimization, resulting in the waveform looking very much like first derivatives of the test functions $\varphi_N(x)$ defined here in (2.1.0), which we use as a building block of the "rotating masks" algorithm. It is interesting to note that in the final detection scheme [Can.] approximates his optimal waveform, derived with the use of statistical estimation, with first derivative of a Gaussian. The two dimensional detector in [Can.] is simply a two-dimensional extension of this ad hoc approximation.

The most important conclusion of the above work is a rigorous demonstration for the need of *multiple widths* of the edge detectors depending on the local estimate of the noise energy, and proof that signal to noise ratio may be significantly improved if *directional operators* are used.

It is very amusing that we come here to the same conclusions just on the basis of *deterministic* model provided by the FTED (2.23). Of course, our considerations are of much more general (applied to the arbitrary order singularities) and geometrical nature.

It should be of fundamental importance to carry statistical analysis of [Can.] in the framework of Numerical Analysis of Singularities. In particular, the *stochastic extension* of the FTED should be stated and proved in order to bring probabilistic realm into the computing of singularities. However affinity of our deterministic conclusions with those of [Can.] gives us confidence that the scheme of the "rotating thin mask" is *naturally* adapted to handle random noise perturbations.

Next we shall briefly account for the fact that realistic image is obtained by the finite sampling in the physical imaging system.

The effects of the sampling procedure are well known. They include errors due to *undersampling*, finite sampling pulse effects, errors of the imperfect reconstruction—collectively known as Aliasing effects.

Ideally we would like to have theory of discrete approximation and accuracy analysis when the essential singularity itself is an object of numerical computation. Specifically the following question should be answered first: given n-dimensional surface distribution, what is its "proper" discrete version, and what is the norm to measure the fit of this numerical approximation? While a few ad hoc attempts to answer this question were made in effect by trying to define "*digital edge*" (e.g. see [Har.]), nobody has asked it in the framework of generalized functions of several variables, which seems to be naturally amenable to such discretization. This and further *discrete singularity* theory is yet to be developed. We shall here only illustrate that inspite of the band limited character of the realistic digital image, the Fundamental Theorem of Edge Detection is still applicable.

Let f(x, y) be an ideal image intensity function. In any realistic imaging system, the actual image function is obtained by first sampling f(x, y) with the *finite pulse width* sampling array, whose sampling rate is chosen to overcome effects of spectral overlap, caused by high frequency components of f(x, y). Clearly, if f(x, y) has essential singularities its Fourier spectrum is of infinite extent.

The final samples of the image are estimated by a spatial integration of the *interpolated* sampled image over discrete grid cells. By making some simple transformations, we obtain the form of this sampled image as follows

$$F(x, y) = f(x, y) * H(x, y) \tag{2.151}$$

where filter H(x, y) absorbs sampling and interpolation procedures.

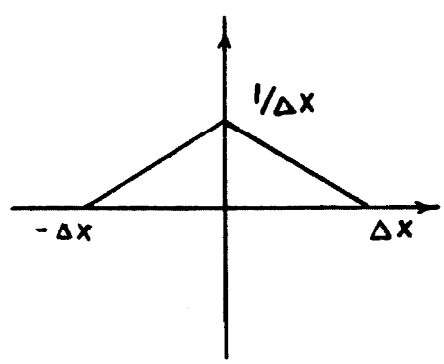
Let Δx, Δy be sampling grid sizes. The corresponding sampling Nyquist rates then are

$$\Delta x = \frac{\pi}{\Omega_x}, \ \Delta y = \frac{\pi}{\Omega_y} \tag{2.152}$$

The resulting F(x, y) is said to be spectrum limited with

$$\text{supp}\{\mathcal{F}(F(x, y))\} = \Omega_x \times \Omega_y \tag{2.153}$$

The bandlimited interpolation to the sampled function is not the only possible. For example, nonbandlimited interpolation is achieved via inscribing linear segments which can be shown to be equivalent to convolving in (2.151) with triangular filter of compact support, i.e:



(Fig. 2.15)

Since H(x, y) in (2.151) is always a smoothing filter, the essential singularities in the ideal image f(x, y) are generally destroyed. Hence one may ask how possibly theory developed here is of any use if singularities are gone? However, a similar situation can be observed in our "rotating masks" algorithm in §2.12 after the first stage of integrating (convolving) image with a sliding smooth mask.

In fact, FTED computes singular contribution *after* the integration stage. Thus, we may use associativity property of convolution and replace the test function $\varphi(x, y)$ in the proof of (2.23) by the new test function

$$\varphi_1(x, y) = \varphi(x, y) * H(x, y) \tag{2.154}$$

where H(x, y) is the imaging system's smoothing filter. In practice H(x, y) is either of small compact support as on Figure 2.15, or it is a rapidly decaying function such as a Gaussian. In either case, the effect of changing test function in (2.154) simply *widens* the support of the detecting mask. However, the whole theory developed here is still applicable without any modifications, except the introduction of the *fundamental limit* on the minimal size of the support of the test function.

As we observed in (2.152) the finite size of the discrete grid imposes Nyquist bounds

$$\Omega_x = \frac{\pi}{\Delta x} \text{ and } \Omega_y = \frac{\pi}{\Delta y}$$

If we use Bernstein's inequality, which establishes fundamental limit on the size of the derivative of bandlimited function of several variables,

$$\left|\frac{\partial^{m+n} f(x, y)}{\partial x^m \partial y^n}\right| \leq \mathrm{Max}(f) \cdot \Omega_x^m \cdot \Omega_y^n, \tag{2.155}$$

the upper bound on the *regular* contribution in the (2.23) may be obtained.

This estimate gives us an upper bound on the error introduced by the sampling and subsequent interpolation of the imaging system.

For example, since the minimal width is one pixel-wide, for the test function $\varphi_1(x, y)$ we have

$$\left\langle \overline{\frac{\partial f}{\partial x}}, \varphi_1 \right\rangle_{reg} \leq \frac{\pi}{\Delta x} \cdot \mathrm{Max}(f) \iint \varphi_1(x, y)\, dxdy \tag{2.156}$$

The use of finite difference operators as in (2.140)–(2.146) can be also rigorously justified since we have Taylor's Theorem for generalized functions (see Theorem 2.7.1). Just as usual Taylor series expansion is used to calculate finite difference approximations to derivatives and evaluate the accuracy of the resulting approximation, the *generalized* Taylor expansion (2.65) will serve similarly to indicate accuracy of the *generalized* finite differences, which (2.140)–(2.146) in effect are.

Finally, we will remark that the formulas for various discontinuous *vector* field quantities are derivable in the manner similar to the FTED. Among them the most important for the application to the analysis of *discontinuous* motion flow are the following identities [Est. Kan.2]:

$$\overline{\mathrm{curl}}(E) = \mathrm{curl}(E) + \overline{n} \times [E] \cdot \delta_L \tag{2.157}$$

$$\overline{\mathrm{div}}(E) = \mathrm{div}(E) + \overline{n} \bullet [E] \cdot \delta_L \tag{2.158}$$

Observe that (2.157)–(2.158) may be perfectly incorporated into our "rotating thin masks" scheme with scalar masks used in (2.157) and vector masks in (2.158). The first of these identities will be useful when expanding singularities are observed in the image, while the second formula should find application in finding boundaries of the rotating bodies in the image.

References

[Asa. Bra.] H. Asado and M. Brady, *The curvature primal sketch*, IEEE transactions on pattern analysis and machine intelligence, Vol. PAMI-8, #1, January 1986, pp. 2–14.

[Ba. Te.] H. Barrow and J. M. Tenenbaum, *Recovering intrinsic scene characteristics from images*, Computer Vision Systems, Ed. A. Hanson, Acad. Pres. 1978.

[Bea.] P. Beaudet, *Rotationally invariant image operators*, 4th Intern. Joint Conf. on Pattern Recognition (1978), Kyoto, Japan.

[Ber.] V. Berzins, *Accuracy of Laplacian edge detectors*, Computer Vision, Graphics, and Image Processing, 27, 195–210, 1984.

[Bin.] T. Binford, *Inferring surfaces from images*, Artificial Intelligence 17(1981), 205–244.

[Bli.] P. Blicher, Edge detection and geometrical methods in computer vision. Report No. STAN.-CS-85-1041.

[Bor. Book] J. Boris and D. Book, *Flux-corrected transport, I. SHASTA, A Fluid Transport Algorithm that Works*, J. Comput. Phys. 11(1973), 38.

[Can.] J. Canny, Finding edges and lines in images, MIT, AI–TR–720, 1983.

[Car.] M. do Carmo, Differential geometry of curves and surfaces, Prentice-Hall, 1976.

[Cour. Hilb] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 2, Interscience Publishers, 1962.

[Est. Kan.1] R. Estrada and R. Kanwal, *Applications of distributional derivatives to wave propagation*, J. Inst. Math. Applics. (1980), 26, pp. 39–63.

[Est. Kan.2] R. Estrada and R. Kanwal, *Distributional analysis for discontinuous fields*, Journal of Mathematical Analysis and Applications, 105, 478–490, 1985.

[Fe.] H. Federer, *Geometric measure theory*, Springer-Verlag, 1969.

[Gel. Shil.] I. M. Gelfand and G. E. Shilov, Generalized functions, Vol. 1, 2.

[Gi.] J. Gibson, *The senses considered as perceptual systems*," 1966.

[Har.] R. Haralick, *Digital step edges from zero crossing of second directional derivatives*, IEEE Trans. on Pattern Analysis, Machine Intelligence, Vol. PAMI-6, No. 1, Jan. 1984.

[Ho.] B.K.P. Horn, *Understanding image intensities*," Artificial Intelligence 8(1977), 201–231.

[Hu.] M. Hueckel, *An operator which locates edges in digital pictures*," JACM, 18, 1, Jan. 1971, 113–125.

[Ke.] J. Kennedy, *The surrogate functions of lines in visual perception: evidence from antipodal rock and cave artwork sources*," Perception, 1974, vol. 3, pp. 313–322.

[K. D.] J. Koendrink and J. van Doorn, *The singularities of the visual mapping*, Biol. Cybernetics 24, 51–59 (1976).

[Kr.] A. Kronrod, *On functions of two variables*, Uspechi Mat. Nank, USSR, (N.S.), 5, 35, 1950, pp. 24–134.

[Lax] P. Lax, *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, SIAM Publish., 1973.

[Lec. Zuc.] Y. Leclerc, S. Zucker, *The local structure of image discontinuities in one dimension*, Proc. International Conf. on Pattern Recognition, Montreal, 1984, pp. 46–48.

[Lev.] S. Levialdi, *Edge extraction techniques*, in Fundamentals in Computer Vision, Ed. O. D. Faugeras, 1983.

[Lu.] M. Luckiesh, Visual Illusions, 1920.

[Ma. Hi.1] D. Marr and E. Hildreth, *Theory of edge detection*, Proc. R. Soc. Lond. B. 207, 187–217 (1980).

[Mar.] D. Marr, Vision, 1982.

[Mars. Tr.] J. Marsden and A. Tromba, Vector Calculus, 1981.

[Mik.] J. Mikusinski, *Irregular operations on distributions*, Studia Mathematica, T. XX (1961).

[Mir.] M. Miranda, *Un teorema di esistenza e unicirá per il problema dell' area minima in n variabili*, Ann. Sc. Norm. Sup. Pisa 19, pp. 233–249 (1965).

[Mor.] C. Morrey, *Multiple integral problems in the calculus of variations and related topics*. Univ. of California Publ. Math. (N. S.) 1, pp. 1–130 (1943).

[Na. Bi.] V. Nalwa and T. Binford, *On detecting edges*," IEEE Tr. PAMI, v.-8, #6, pp. 699–714.

[Pavl.] T. Pavlidis, Algorithms for graphics and image processing, Computer Science Press, 1982.

[Pe. Ro.] D. Pearson and J. Robinson, *Visual communication at very low data rates*," Proc. IEEE, Vol. 73, no. 4, pp. 795–812.

[Pra.] W. Pratt, *Digital Image Processing*, J. Wiley & Sons, 1978.

[P.F.T.V.] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, *Numerical Recipes*, Cambridge University Press.

[Ric. Mor.] R. Richtmyer and K. Morton, Difference methods for initial-value problems. Interscience publishers, 1967.

[Ros. Kak.] A. Rosenfeld, A. Kak, Digital Picture Processing, Vol. 2, Academic Press 1982.

[Rud.1] L. Rudin, *Shock Filters*, DARPA Report, Rockwell International, Science Center, 1984.

[Rud.2] L. Rudin, *Images and Shock Filters*, (to be published).

[Sha. Jai.] M. Shah and R. Jain, *Detecting Time-varying Corners*, IEEE transactions on pattern analysis and machine intelligence, VOL. PAMI VII, 1984, pp. 2–14.

[Ve. Ko.1] I. Verchenko and A. Kolmogorov, *On points of discontinuity of functions of two variables*, Doklady Academy Sc. USSR #3, pp. 105–105, 1934.

[Ve. Ko.2] *Continuation of investigation of points of discontinuity of functions of two variables*, Doklady Acad. Sc. USSR, IV #7, 361–362.

[Vol.] A. I. Volpert, *The spaces BV and quasilinear equations*, Math. USSR-Sbornik, Tom 73(115), 1967, #2, pp. 225–267.

[Wa.] D. Waltz, *Constraint propagation in interpreting line drawings, The Psychology of computer vision* by P. Watson,

[Wh.] H. Whitney, *On singularities of mappings of Euclidean spaces*, Annals of Mathematics, Vol. 62, #3, 1955, pp. 74–410.

[Wit.] A. Witkin, *Scale space filtering*, Proceed. of Intern. Joint Conf. on Artificial Intelligence, Karlsruhe 1983, pp. 1019–1021.